Einführung in die Stochastik für Informatiker
Vorlesungsmitschrift SS 2000
Prof. Mathar

geLaTeXt von

David Buttgereit
david.buttgereit@post.rwth-aachen.de

14. Juli 2000

Inhaltsverzeichnis

1 Einleitung

Betrachte "Zufallsexperimente", z.B.

- Münzwurf, Würfelwurf, Spiele, Roulette, Lotto
- Ankunft von Kunden an Schaltern, Pakete in Netzwerken
- Input für Algorithmen
- Signale, die von einer Quelle ausgesendet werden
- Positionierung von Mobilstationen in Zellnetzen

Gemeinsam in diesen Beispielen: Interessierende Größen können nicht genau vorhergesagt werden, sind zufallsabhängig
Stochastik: mathematische Behandlung von Zufallsphänomenen <u>Teilgebiete:</u>

- Wahrscheinlichkeitstheorie
  - theoretisch (stochastische Prozesse, Grenzwertsätze, stochastische Diffentialgleichungen)
  - angewandt = stochastische Modellierung (Warteschlangen, Zuverlässigkeitstheorie, stochastische Signalerkennung)
- mathematische Statistik (mit Teilgebieten)

Hier: Wahrscheinlichkeitstheorie, angewandt, stochastische Modellierung, Betonung der Anwendung in der Informatik
<u>Ziel der Vorlesung:</u> Bereitstellung der Grundlagen, Basis für weiterführende Vorlesungen

> Warteschlangensysteme (I), Warteschlangennetze (II), Informationstheorie I,II, Kryptologie, stochastische Simulation, zufallsgesteuerte Optimierungsverfahren

Historiche Entwicklung

> Fermat (1601-65), Pascal (1623-62), Bernoulli (1654-1705)
> Laplace (1749-1827): Kombinatorischer Zugang, motiviert durch Spielprobleme, relative Häufigkeiten als Wahrscheinlichkeiten
> Kolmogoroff (1933): Axiomatische Entwicklung

2 σ-Algebren und W'keitsverteilungen

Mathematische Beschreibung von Zufallsexperimenten mit Mengen.
Betrachte relevante **Ergebnisse** und fasse sie zu einer Menge zusammen.
Ω: **Ergebnismenge** (oft $\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{R}^n, \{0,1\}^{\mathbb{N}}$)
Ereignisse werden durch Teilmengen von Ω beschrieben.
Zunächst $\mathcal{P}(\Omega)$ als Ereignismenge.

W'keiten von Ereignissen durch Funktion $P : \mathcal{P}(\Omega) \to [0,1]$ mit den Eigenschaften (*):
$P(\Omega) = 1$, $P(\emptyset) = 0$, $P(A^C) = 1 - P(A) \quad \forall A \in \mathcal{P}(\Omega)$,
$P(A \cup B) = P(A) + P(B) \quad \forall A, B \in \mathcal{P}(\Omega)$ mit $A \cap B = \emptyset$,

σ-Additivität:
$P(\bigcup_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} P(A_i) \quad A_i \in \mathcal{P}(\Omega)$ mit A_i paarweise disjunkt [1]

Schrechweisen:
A^C: "A tritt nicht ein", $A \in \mathcal{P}(\Omega)$
$A \cup B$: "A oder B treten ein", $A, B \in \mathcal{P}(\Omega)$
$A \cap B$: "A und B treten ein", $A, B \in \mathcal{P}(\Omega)$

Wie erhält man W'keiten? Bei endlichem Ω durch Abzählen.

Definition 2.1 (Laplacescher W'begriff)
Ω sei eine endliche Menge. $|A|$ bezeichnet die Mächtigkeit von $A \subseteq \Omega$. Durch $P(A) = \frac{|A|}{|\Omega|}$ für $A \in \mathcal{P}(\Omega)$ wird eine W'verteilung P auf $\mathcal{P}(\Omega)$ mit den Eigenschaften (*) definiert. P heißt **Laplace-Verteilung** oder **diskrete Gleichverteilung** über Ω.

Beispiel 2.2 (Binäre Suche)
Geg. geordnetes Feld von $2^n - 1$ Elementen.
Schlüsselelement y vergleichbar mit den Elementen.
Problem: Ist y in dem Feld vorhanden und an welcher Stelle?
Algorithmus: Binary search
Stochastisches Modell:
$\Omega = \{0, 1, \ldots, 2^n - 1\}$, $w \in \Omega = \begin{cases} w \geq 1: & \text{gesuchtes Element an Stelle } w \\ w = 0: & y \text{ nicht vorhanden} \end{cases}$
Bestimme Ereignisse $A_k : y$ wird im k-ten Schritt gefunden. $A_1 = \{2^{n-1}\}$,
$A_2 = \{2^{n-2}, 3 \cdot 2^{n-2}\}, \ldots, A_k = \{(2j-1)2^{n-k} | j = 1, \ldots, 2^{k-1}\}$ für $k = 1, \ldots, n$
Annahme: Jede Platznummer und 0 sind gleichwahrscheinlich. Dann $|A_k| = 2^{k-1}$, $P(A_k) = \frac{2^{k-1}}{2^n}$ für $1 \leq k \leq n \Longrightarrow$ W'keit, y in genau k Schritten zu finden

Zusammengesetzte Ereignisse $B_k : y$ wird in höchstens k Schritten gefunden

[1] d.h. $A_i \cap A_j = \emptyset \quad \forall i \neq j$

$B_k = \sum_{j=1}^{k} A_j$ mit $j = 1, \ldots, k$ und die A_j disjunkt

$P(B_k) = P(\bigcup_{j=1}^k A_j) = \sum_{j=1}^{k} P(A_j) = \sum_{j=1}^{k} \frac{2^{j-1}}{2^n} = \frac{1}{2^n}(2^k - 1) = \frac{2^k-1}{2^n}$

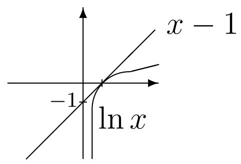
Komplizierte Anzahlbestimmungen durch kombinatorische Überlegungen.

Beispiel 2.3 (Hashing)
Universum U, $M \subseteq U$, $|M| = k$, M soll abgespeichert werden
Hashtafel: $a : array[0, \ldots, n-1]$ of ...
Hashfunktion: $h : U \rightarrow [0, \ldots, n-1]$ Einfache Funktion!
Speichere $x \in M$ in $a[h(x)]$
Kollision, falls $h(x) = h(y)$ für $x \neq y$ (Auflösung durch linerare Liste)

Stochastisches Modell:
(”Rein zufälliges” Ablegen von k Daten in Feld der Länge n)
$S = \{1, \ldots, n\}$ (Speicherplätze)
$\Omega = S^k$ ($k-n-$Permutationen mit Wiederholungen/Kollisionen)
$A_{k,n}$ ($k-n-$Permutationen ohne Wiederholungen/Kollisionen)

Wahrscheinlichkeit für kollisionsfreie Abspeicherung:
$P(A_{k,n}) = \frac{|A_{k,n}|}{|\Omega|} = \frac{n(n-1)\ldots(n-k+1)}{n^k} = \prod_{i=0}^{k-1}(1-\frac{i}{n}) = \exp(\ln(\prod_{i=0}^{k-1}(1-\frac{i}{n})))$
$= \exp(\sum_{i=0}^{k-1} \ln(1-\frac{i}{n}))$
$\overset{(*)}{\leq} \exp(-\sum_{i=0}^{k-1} \frac{i}{n})$
$= \exp(-\frac{(k-1)k}{2n})$

Benutzte Abschätzung (∗):
$\ln x \leq x - 1 \quad \forall x \geq 0$
$\Longrightarrow \ln(1-x) \leq -x \quad \forall x \leq 1$



Zahlenbeispiel: Für $n = 365$ und $k = 23$ ist $P(A_{k,n}) \leq 0,499998$

Der Laplacesche W'begriff reicht nicht aus, da

- Ω ist oft ∞ oder überabzählbar
- Viele Experimente sind nicht durch eine diskrete Gleichverteilung beschreibbar

Beispiel 2.4 (∞-Münzwurf) Kopf$\widehat{=}$1, Zahl$\widehat{=}$0
Bestimme ein Modell für nicht-abbrechende Folge von Münzwürfen (wird gebraucht bei der Frage ”Wann tritt zum ersten Mal Kopf (Zahl) auf?”. Jeder beliebig ferne Wurf könnte der erste sein.).

Stochastisches Modell:
$\Omega = \{w = (x_1, x_2, \ldots) | x_i \in \{0, 1\}\} = \{0, 1\}^{\mathbb{N}}$ Menge der 0-1-Folgen
Bekannt: Ω ist überabzählbar.

Problem: Beschreibe Gleichverteilung auf Ω Ansatz: $P(\{w\}) = \delta > 0$
Sei $A = \{w_1, w_2, \ldots\} \subset \Omega$ unendlich, aber abzählbar. Dann $P(A) = \sum\limits_{i=1}^{\infty} P(\{w_i\})$
$= \sum\limits_{i=1}^{\infty} \delta = \infty$ Wiederspruch zu $P(\Omega) = 1$.
Also: $P(\{w\}) = 0 \quad \forall w \in \Omega$ Wenig hilfreich!

Beispiel 2.5 (Gleichverteilung über $[0,1]$**)**
(z.B. jede reelle Zahl zwischen 0 und 1 ist gleich wahrscheinlich)
$\Omega = [0,1]$ (Ergebnismenge)
Versuch: Ereignismenge $= \mathcal{P}(\Omega)$
Forderungen:

1. $P([a,b]) = b - a \quad \forall 0 \leq a < b \leq 1$
2. P ist σ-additiv

Es gilt: Eine Funktion P mit diesen Eigenschaften existiert nicht. Im $\mathbb{R}^3$ gibt es sogar keine, wenn nur endliche Additivität verlangt wird: Es gibt kein endlich additives drehinvariantes[2] Maß auf $\mathcal{P}(\mathbb{R}^3)$. (Hausdorff 1914)

Für einen allgemeinen W'begriff (ohne Existenzprobleme):
Ereignismenge nicht $\mathcal{P}(\Omega)$, sondern kleineres Mengensystem wählen, das noch alle interessanten Ereignisse enthält.

Definition 2.6 Sei $\Omega \neq \emptyset$, $\mathfrak{A} \subseteq \mathcal{P}(\Omega)$ ein System von Teilmengen von Ω. $\mathfrak{A}$ heißt **σ-Algebra** (von Ereignissen) über Ω, wenn

(i) $\Omega \in \mathfrak{A}$
(ii) $A \in \mathfrak{A} \Longrightarrow A^C \in \mathfrak{A}$
(iii) $A_n \in \mathfrak{A} \quad \forall n \in \mathbb{N} \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$

$(\Omega, \mathfrak{A})$ heißt **Meßraum**.

Mit den De Morgan-Regeln: $A_n \in \mathfrak{A} \overset{ii)}{\Longrightarrow} A_n^C \in \mathfrak{A} \overset{iii)}{\Longrightarrow} \bigcup_{n=1}^{\infty} A_n^C \in \mathfrak{A} \overset{ii)}{\Longrightarrow}$
$(\bigcup_{n=1}^{\infty} A_n^C)^C \in \mathfrak{A} \Longrightarrow \bigcap_{n=1}^{\infty} A_n \in \mathfrak{A} \quad \forall n \in \mathbb{N}$
σ-Algebren enthalten alle Ereignisse, die durch Verknüpfungen (abzählbar vieler Ereignisse) mit "und", "oder" und "nicht" entstehen. Dies ist wichtig für die Festlegung von W'verteilungen.

Beispiel 2.7

a) $\mathcal{P}(\Omega)$ ist stets eine σ-Algebra (die feinste σ-Algebra)

[2] Wenn man im $\mathbb{R}^3$ einen Körper dreht, muss er derselbe bleiben.

b) $\{\emptyset, \Omega\}$ ist eine σ-Algebra (die gröbste σ-Algebra)

c) $\Omega = \mathbb{N}$, $G = \{2, 4, 6, \ldots\}$, $U = \{1, 3, 5, \ldots\}$, $\mathfrak{A} = \{\emptyset, G, U, \Omega\}$ ist σ-Algebra

d) $\Omega = \mathbb{R}$, $E = \{(a, b] | a < b \text{ mit } a, b \in \mathbb{R}\}$ ist keine σ-Algebra
Denn sei $a < b < c < d$. $(a, b] \in E$ und $(c, d] \in E$, aber $(a, b] \cup (c, d] \notin E$.

Problem: Gibt es eine kleinste (im Sinne der Mengeninklusion) σ-Algebra, die E enthält?

Lemma 2.8 $\Omega \neq \emptyset$, $\mathfrak{A}_i$ sei σ-Algebra über Ω für alle $i \in I$. Dann ist $\bigcap_{i \in I} \mathfrak{A}_i$ ebenfalls eine σ-Algebra. (Beweis in der Übung 2)

Sei $E \subseteq \mathcal{P}(\Omega)$. $\mathfrak{A}(E) = \bigcap\limits_{\substack{\mathfrak{A} \text{ ist } \sigma\text{-Algebra} \\ \text{und } E \subseteq \mathfrak{A}}} \mathfrak{A}$ ist die kleinste σ-Algebra, die E enthält, ist die **von E erzeugte σ-Algebra**.

Definition 2.9 Sei $\Omega = \mathbb{R}$, $E = \{(a, b] | a < b \text{ mit } a, b \in \mathbb{R}\}$. $\mathcal{B} = \mathfrak{A}(E)$ heißt **Borelsche σ-Algebra**.

Auf σ-Algebren können W'verteilungen mit den Eigenschaften (*) (Seite 4) definiert werden.

Definition 2.10 $\Omega \neq \emptyset$, $\mathfrak{A}$ eine σ-Algebra. Eine Abbildung $P : \mathfrak{A} \to [0, 1]$ mit

(i) $P(\Omega) = 1$

(ii) $P(\bigcup_{n=1}^{\infty} A_n) = \sum\limits_{n=1}^{\infty} P(A_n) \quad \forall A_n \in \mathfrak{A}$, paarweise disjunkt

heißt **Wahrscheinlichkeitsverteilung** oder **Wahrscheinlichkeitsmaß** auf $(\Omega, \mathfrak{A})$. $(\Omega, \mathfrak{A}, P)$ heißt **Wahrscheinlichkeitsraum**.

Der Laplacesche W'begriff (Def. 2.1) ist ein Spezialfall von Def. 2.10. Denn für $\mathfrak{A} = \mathcal{P}(\Omega)$ erfüllt $P(A) = \frac{|A|}{|\Omega|}$

(i) $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$

(ii) $P(\bigcup_{n=1}^{\infty} A_n) = \frac{|\bigcup_{n=1}^{\infty} A_n|}{|\Omega|} = \frac{\sum\limits_{n=1}^{\infty} |A_n|}{|\Omega|} = \sum\limits_{n=1}^{\infty} \frac{|A_n|}{|\Omega|} = \sum\limits_{n=1}^{\infty} P(A_n)$
$\forall A_n$ paarweise disjunkt

Sei $(\Omega, \mathfrak{A}, P)$ W'raum. Eine Folge von Ereignissen $\{A_n\}_{n \in \mathbb{N}}$ heißt **aufsteigend** (**absteigend**), wenn $A_n \subseteq A_{n+1}$ ($A_{n+1} \subseteq A_n$) für alle $n \in \mathbb{N}$. $\bigcup_{n=1}^{\infty} A_n = \lim\limits_{n \to \infty} A_n$ ($\bigcap_{n=1}^{\infty} A_n = \lim\limits_{n \to \infty} A_n$) heißt **Limes** der Mengenfolge $\{A_n\}$.

Lemma 2.11 (Eigenschaften von W'verteilungen)
$(\Omega, \mathfrak{A}, , P)$ sei W'raum. $A, B, A_n \in \mathfrak{A} \quad \forall n \in \mathbb{N}$.

a) $P(A^C) = 1 - P(A)$
b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
c) $P(A) \leq P(B)$ falls $A \subseteq B$
d) $P(B \setminus A) = P(B) - P(A)$ falls $A \subseteq B$
e) $\{A_n\}$ aufsteigend $\Longrightarrow P(\lim\limits_{n\to\infty} A_n) = \lim\limits_{n\to\infty} P(A_n)$ (Stetigkeit von unten)
$\{A_n\}$ absteigend $\Longrightarrow P(\lim\limits_{n\to\infty} A_n) = \lim\limits_{n\to\infty} P(A_n)$ (Stetigkeit von oben)

Beweis: a)-d) in der 2. Übung
e) 1.) $\{A_n\}$ aufsteigend, d.h. $A_n \subseteq A_{n+1} \quad \forall n \in \mathbb{N}$
Setze $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, ..., $B_{n+1} = A_{n+1} \setminus A_n$
Dann gilt B_n p.d. (paarweise disjunkt) und $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n$. Setze $A_0 = \emptyset$,
dann ist $P(\bigcup_{n=1}^\infty A_n) = P(\bigcup_{n=1}^\infty B_n) \overset{ii)}{=} \sum\limits_{n=1}^{\infty} P(B_n) = \lim\limits_{k\to\infty} \sum\limits_{n=1}^{k} P(A_n \setminus A_{n-1})$
$= \lim\limits_{k\to\infty} \sum\limits_{n=1}^{k} [P(A_n) - P(A_{n-1})] = \lim\limits_{k\to\infty} P(A_k)$
2.) $\{A_n\}$ absteigend $\Longrightarrow \{A_n^C\}$ aufsteigend. Es gilt:
$P(\bigcap_{n=1}^\infty A_n) = 1 - P(\bigcup_{n=1}^\infty A_n^C) \overset{1.)}{=} 1 - \lim\limits_{n\to\infty} P(A_n^C) = \lim\limits_{n\to\infty} (1 - P(A_n^C))$
$= \lim\limits_{n\to\infty} P(A_n)$ □

Lemma 2.12 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $A_1, \ldots, A_n \in \mathfrak{A}$.

a) **Siebformel** von Pomcare-Sylvester (inclusion-exclusion-formula)
$P(\bigcup_{i=1}^n A_i) = \sum\limits_{i=1}^{n} P(A_i) - \sum\limits_{1\leq i_1<i_2\leq n} P(A_{i_1} \cap A_{i_2}) + \sum\limits_{1\leq i_1<i_2<i_3\leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \ldots + \ldots + (-1)^{n+1} P(\bigcap_{i=1}^n A_i)$
b) **Bonferroni-Ungleichung** (Bonferroni inequality)
$\sum\limits_{i=1}^{n} P(A_i) - \sum\limits_{1\leq i_1<i_2\leq n} P(A_{i_1} \cap A_{i_2}) \leq P(\bigcup_{i=1}^n A_i) \leq \sum\limits_{i=1}^{n} P(A_i)$
Weitere obere bzw. untere Schranken durch Abbruch in a) nach + oder - Zeichen

Beispiel 2.13 (Sortieren) (Recourtre-Problem)
Betrachte Feld der Länge n von vergleichbaren, verschiedenen Elementen. Alle Anordnungen seien gleichwahrscheinlich.
Stoch. Modell: $\Omega = \{\text{Permutationen von } \{1, \ldots, n\}\}$, $\mathfrak{A} = \mathcal{P}(\Omega)$
$P(\{w\}) = \frac{1}{|\Omega|} = \frac{1}{n!} \quad \forall w \in \Omega$

a) Bestimme die W'keit, daß mindestens ein Element an der richtigen Stelle ist (vorsortiert).

$A_j = \{w = (w_1, \ldots, w_n) \in \Omega | w_j = j\}$ j-tes Element an der richtigen Stelle
Gesucht: $P(A_1 \cup \ldots \cup A_n) = P(\bigcup_{i=1}^n A_i)$
Berechnung mit Hilfe der Siebformel: $1 \leq i_1 < i_2 \leq n$, $l \leq n$
$\bigcap_{j=1}^l A_{i_j} = \{w \in \Omega | w_{i_j} = i_j,\ j = 1, \ldots, l\}$, $|\bigcap_{j=1}^l A_{i_j}| = (n-l)!$
Also $P(\bigcap_{j=1}^l A_{i_j}) = \frac{(n-l)!}{n!} = \frac{1}{\binom{n}{l} l!}$ mit $l = 1, \ldots, n$
Außerdem: $|\{(i_1, \ldots, i_l) | 1 \leq i_1 < \ldots < i_l \leq n\}| = \binom{n}{l}$
Insgesamt: $P(\bigcup_{j=1}^n A_j) = \sum_{j=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \ldots + (-1)^{n+1} P(\bigcap_{j=1}^{} A_j)$
$= \frac{n}{n} - \binom{n}{2} \frac{1}{\binom{n}{2} 2!} + \binom{n}{3} \frac{1}{\binom{n}{3} 3!} - \ldots + (-1)^{n+1} \binom{n}{n} \frac{1}{\binom{n}{n} n!}$
$= 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + \frac{(-1)^{n+1}}{n!}$
$= 1 - (1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots + \frac{(-1)^n}{n!}) \stackrel{n \to \infty}{\to} 1 - e^{-1} \approx 0,6321$
n groß: W'keit für mindestens ein vorsortiertes Element ist $0,6321$ (fast unabhängig von n)

b) Bestimme die W'keit, daß mindestens k Elemente vorsortiert sind.
$$P(\bigcup_{1 \leq i_i < \ldots < i_k \leq n} \bigcap_{l=1}^k A_{i_l}) \stackrel{\textit{Lemma } 2.12b)}{\leq} \sum_{1 \leq i_i < \ldots < i_k \leq n} P(\bigcap_{l=1}^k A_{i_l}) = \binom{n}{k} \frac{1}{\binom{n}{k} k!} = \frac{1}{k!}$$
fällt sehr schnell mit steigendem k

c) Bestimme die W'keit, daß genau k Elemente vorsortiert sind.
Nach a): W'keit, daß in einem Feld der Länge $n-k$ kein Element vorsortiert ist: $1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + \frac{(-1)^n}{n!}$
$\Longrightarrow$ Anzahl der Anordnungen, beim denen kein Element vorsortiert ist: $(n-k)!(1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + \frac{(-1)^n}{n!})$
Anzahl der Möglichkeiten, ein Feld der Länge n in ein Teilfeld der Länge $n-k$ und eines der Länge k aufzuteilen: $\binom{n}{k}$
Insgesamt: W'keit, daß genau k Elemente vorsortiert sind:
$\frac{1}{n!} \binom{n}{k} (n-k)!(1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + \frac{(-1)^n}{n!}) = \frac{1}{k!}(1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + \frac{(-1)^{n-k}}{(n-k)!})$

Wie legt man W'keiten fest, wenn schon bekannt ist, daß das Ergebnis in einer bestimmten Teilmenge liegt?

Motivierendes Beispiel 2.14 5000 Chips von zwei Firmen. Ω mit $|\Omega| = 5000$.
A: Chips der Firma A, mit $|A| = 1000$
B: Chips der Firma B, mit $|B| = 4000$
D: Defekte Chips, mit $|D| = 300$
Gelte: $|A \cap D| = 100$ (10% defekt) und $|B \cap D| = 200$ (5% defekt)
Ziehe zufällig einen Chip (Laplace-Modell).
W'keit, daß der Chip defekt ist, wenn er von Firma A stammt:
$P(D|A) = \frac{|D \cap A|}{|A|} = \frac{|D \cap A|/\Omega}{|A|/\Omega} = \frac{P(D \cap A)}{P(A)} = \frac{100/5000}{1000/5000} = 0,1$
Auch: $P(A|D) = \frac{|A \cap D|}{|D|} = \frac{|A \cap D|/\Omega}{|D|/\Omega} = \frac{P(A \cap D)}{P(D)} = \frac{100/5000}{300/5000} = \frac{1}{3}$

Definition 2.15 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $A, B \in \mathfrak{A}$, $P(B) > 0$. $P(A|B) = \frac{P(A \cap B)}{P(B)}$ heißt **(elementare) bedingte Wahrscheinlichkeit von** A **unter (der Hypothese)** B. Durch $P(A^3|B) : \mathfrak{A} \to [0,1]$ mit $A \mapsto P(A|B)$ wird eine W'verteilung auf $\mathfrak{A}$ definiert, die (elementare) **bedingte Verteilung unter** B.

Satz 2.16 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $B_n \in \mathfrak{A}$ für $n \in \mathbb{N}$ sei Partition von Ω, d.h. $\bigcup_{n=1}^{\infty} B_n = \Omega$ und B_n paarweise disjunkt.

a) **Satz von der totalen W'keit**
Für alle $A \in \mathfrak{A}$ gilt:
$P(A) = \sum_{n=1}^{\infty} P(A|B_n)P(B_n)$ (Konvention: "Nicht definiert"$\cdot 0 = 0$, z.B. $\frac{x}{0} \cdot 0 = 0$)

b) **Bayes-Formel**
Falls $P(A) > 0$, gilt für alle $n \in \mathbb{N}$
$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{\sum_{j=1}^{\infty} P(A|B_j)P(B_j)}$

Beweis:

a) $P(A) = P(\dot{\bigcup}_{n=1}^{\infty} A \cap B_n) = \sum_{n=1}^{\infty} P(A \cap B_n) = \sum_{n=1}^{\infty} P(A|B_n)P(B_n)$ $\square$

b) $P(B_n|A) = \frac{P(B_n \cap A)}{P(A)} = \frac{P(A|B_n)P(B_n)}{\sum_{j=1}^{\infty} P(A|B_j)P(B_j)}$ $\square$

Wichtiger Spezialfall: $P(A|B) = P(A|B^C)$, wobei $P(B) > 0$, $P(B^C) > 0$ $(*)$
(W'keit für das Eintreten von A hängt nicht vom Eintreten von B ab.)
Dann gilt: $P(A) = P(A \cap B) + P(A \cap B^C) = P(A|B)P(B) + P(A|B^C)P(B^C) \overset{(*)}{=} P(A|B)(P(B) + P(B^C)) = P(A|B) = \frac{P(A \cap B)}{P(B)}$
Also: $P(A) = \frac{P(A \cap B)}{P(B)} \iff P(A)P(B) = P(A \cap B)$
Diese Definition wird erweitert auf n Ereignisse $A_1, \ldots, A_n \in \mathfrak{A}$, bzw. auf Folgen von Ereignissen.

Definition 2.17 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $A_1, \ldots, A_n \in \mathfrak{A}$ Ereignisse. $A_1, \ldots, A_n$ heißen **(gemeinsam) stochastisch unabhängig** , wenn $P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k})$ für alle $1 \leq i_1 < \ldots < i_k \leq n$ und für alle k.
Eine Folge $\{A_n\}_{n \in \mathbb{N}}$ von Ereignissen heißt stochastisch unabhängig, wenn die $A_1, \ldots, A_n$ für alle $n \in \mathbb{N}$ stochastisch unabhängig sind.

Beachte: Aus paarweiser stochastischer Unabhängigkeit folgt nicht die (gemeinsame) stochastische Unabhängigkeit!

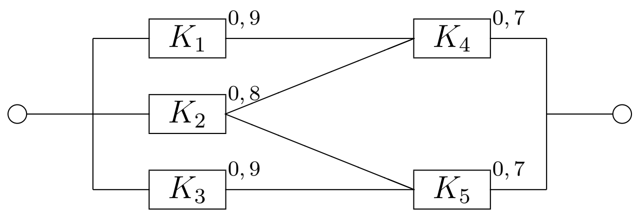
Lemma 2.18 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $A_1, \ldots, A_n \in \mathfrak{A}$.
Es gilt: $A_1, \ldots, A_n \in \mathfrak{A}$ stochastisch unabhängig $\iff \forall B_i \in \{A_i, A_i^C\}$, $i = 1, \ldots, n$: $P(\bigcap_{i=1}^{n} B_i) = \prod_{i=1}^{n} P(B_i)$

[3] A variabel in $\mathfrak{A}$

Es gilt: $A_1, \ldots, A_n \in \mathfrak{A}$ stochastisch unabhängig $\Longleftrightarrow \forall B_i \in \{A_i, A_i^C\}$, $i = 1, \ldots, n$: $B_1, \ldots, B_n$ stochastisch unabhängig

Beweis: Stochastik für Informatiker (Mathar & Pfeiffer)

Beispiel 2.19 System bestehe aus 5 Komponenten:



System ist intakt, wenn mindestens ein "Pfad" intakt ist.
W'keit, daß Komponente i intakt ist: p_i für $i = 1, \ldots, 5$ (Werte in Skizze)
$\{$Komponente K_i intakt$\}$ seien stochastisch unabhängige Ereignisse.

Modellbildung: $\Omega = \{(x_1, \ldots, x_5) | x_i \in \{0,1\}\}$ ($x_i = 1$: Komponente i intakt)
$A_i = \{(x_1, \ldots, x_5) | x_i = 1\}$, $P(A_i) = p_i$ (Komponente i intakt)
$A_1, \ldots, A_5$ stochastisch unabhängig
$S = (A_1 \cap A_4) \cup (A_2 \cap A_4) \cup (A_2 \cap A_5) \cup (A_3 \cap A_5)$ (System intakt)
$P(S) = P((A_1 \cap A_4) \cup (A_2 \cap A_4) \cup (A_2 \cap A_5) \cup (A_3 \cap A_5))$ Vereinigung nicht disjunkter Mengen $\Longrightarrow$ Berechnung mit Sylvesterformel möglich, aber sehr aufwendig!

Mit Satz 2.15 a):
$$P(S) = \underbrace{P(S|A_2)P(A_2)}_{(*1)} + \underbrace{P(S|A_2^C)}_{(*2)} P(A_2^C)^4$$
$$(*1): P(S|A_2) = P(A_4 \cup A_5) = 1 - P(A_4^C \cap A_5^C) = 1 - P(A_4^C)P(A_5^C)$$
$$= 1 - (1 - p_4)(1 - p_5)$$
$$(*2): P(S|A_2^C) = P((A_1 \cap A_4) \cup (A_3 \cap A_5)) = 1 - P((A_1 \cap A_4)^C \cap (A_3 \cap A_5)^C)$$
$$= 1 - (1 - p_1 p_4)(1 - p_3 p_5)$$
Insgesamt: $P(S) = [1 - (1 - p_4)(1 - p_5)]p_2 + [1 - (1 - p_1 p_4)(1 - p_3 p_5)](1 - p_2)$
$= \ldots = 1 - p_2(1 - p_4)(1 - p_5) - (1 - p_2)(1 - p_1 p_4)(1 - p_3 p_5) = 0{,}90062$

Mit Satz 2.15 b): (W'keit, daß Komponente 2 intakt, wenn System intakt)
$P(A_2|S) = \frac{P(S|A_2)P(A_2)}{P(S)} = \frac{0{,}91 \cdot 0{,}8}{0{,}90062} = 0{,}80833$

Betrachte Limites von Mengenfolgen die nicht notwendig auf- oder absteigend sind.

[4] A_2 und A_2^c stellen eine Mengenpartition von Ω dar, also Satz der totalen W'Keit

Beispiel 2.20 (unendlicher Münzwurf)
$\Omega = \{w = (x_1, x_2, \ldots)|x_i \in \{0, 1\}\}$
$A_n = \{w = (x_1, x_2, \ldots)|x_n = 1\}$ (im n-ten Wurf fällt Kopf)
Beschreibe das Ereignis A: es fällt ∞-oft Kopf = es treten ∞ viele der A_n ein
$A = \{w|w \in A_n$ für ∞ viele $n\} = \{w|\forall k \exists n \geq k : w \in A_n\} = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty}$
Analog: B: Fast alle der A_n treten ein (alle bis auf endlich viele Ausnahmen)
$B = \{w|w \in A_n$ für fast alle $n\} = \{w|\exists k \forall n \geq k : w \in A_n\} = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty}$

Definition 2.21 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $A_n \in \mathfrak{A}$, $n \in \mathbb{N}$

$\limsup\limits_{n\to\infty} A_n = \bigcap\limits_{k=1}^{\infty} \bigcup\limits_{n=k}^{\infty} A_n$ heißt **Limes superior** der $\{A_n\}$

$\liminf\limits_{n\to\infty} A_n = \bigcup\limits_{k=1}^{\infty} \bigcap\limits_{n=k}^{\infty} A_n$ heißt **Limes inferior** der $\{A_n\}$

Es ist:

$$\limsup_{n\to\infty} A_n = \lim_{n\to\infty} \left(\bigcup_{k=n}^{\infty} A_k \right) = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

$$\liminf_{n\to\infty} A_n = \lim_{n\to\infty} \left(\bigcap_{k=n}^{\infty} A_k \right) = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

3 Zufallsvariable und ihre Verteilung

Zur Motivation: Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$: Zur Beschreibung von Zufallsexperimenten, zur Modellierung von Zufallseinflüssen.

- Oft interessiert nicht das gesamte Modell, sondern nur Teilgrößen.
  Z.B. 5000 Stichproben mit Ausgang gut/schlecht. Von Interesse sind nun aber z.B. die Anzahl der schlechten Stücke.
  $\Omega = \{(x_1, ..., x_{5000}) | x_i \in \{g, s\}\}$ ursprüngliche Ergebnismenge
  $T = \{0, ..., 5000\}$ von Ereignissen

- Ergebnisse von Zufallsexperimenten sind oft Zahlen oder Vektoren. Nützlich oft: arithmetische Operationen.
  Z.B.: 5000 Münzwürfe mit Ergebnis 0 oder 1 (s.o.) $x_1, ...x_{5000} \in \{0, 1\}$
  Von Interesse: Anzahl der 1'en $= \sum\limits_{i=1}^{5000} x_i$
  oder
  $x_1, ..., x_n \in \mathbb{R}^+$ (Z.B. Verzögerungszeiten im Switch)
  $\overline{x} = \frac{1}{n} \sum\limits_{i=1}^{n} x_i$ (mittlere Verzögerungszeit, Mittelwert)

Modellierung allgemein mit Zufallsvariablen.

Definition 3.1 (Zufallsvariable)
$(\Omega, \mathfrak{A}, P)$ sei W'Raum. Eine Abb. $X : \Omega \to \mathbb{R}$ heißt **Zufallsvariable** (ZV) oder **Zufallsgröße** (random variable, r.v.), wenn $X^{-1}(B) = \{\omega \in \Omega | X(\omega) \in B\} = \{X \in B\} \in \mathfrak{A} \quad \forall B \in \mathcal{B}^1$
Diese Bedingung heißt **Meßbarkeit von X**
Schreibweise: $X : (\Omega, \mathfrak{A}) \to (\mathbb{R}, \mathcal{B}^1)$ **(X ist Abb. und meßbar)**

Bemerkung: Der Begriff 'Zufallsvariable' hat sich eingebürgert, obwohl es sich eigentlich um eine 'Funktion' handelt. Betont wird die Modellierung des Zufalls, wichtig ist die Verteilung von Zufallsvariablen.

Lemma 3.2 $X : (\Omega, \mathfrak{A}) \to (\mathbb{R}, \mathcal{B}^1)$ sei ZV'e. durch $P^X(B) := P(X^{-1}(B)) = P(\{\omega | X(\omega) \in B\}) = P(X \in B)$, $B \in \mathcal{B}^1$ wird eine W'verteilung auf $(\mathbb{R}, \mathcal{B}^1)$ definiert.
P^X heißt **Verteilung der ZV'en X**.

Beweis: Überprüfe Def. 2.10

(i) $P^X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$

(ii) B_n paarweise disjunkt, $B_n \in \mathcal{B}^1$, $n \in \mathbb{N}$

$$P^X(\bigcup_{n=1}^\infty B_n) = P(\underbrace{X^{-1}\underbrace{(\bigcup_{n=1}^{\infty} B_n)}_{\in\mathcal{B}^1}}_{\in\mathfrak{A}}) = P(\bigcup_{n=1}^\infty \underbrace{X^{-1}(B_n)}_{p.w.disjunkt})$$

$$= \sum_{n=1}^{\infty} P(X^{-1}(B_n)) = \sum_{n=1}^{\infty} P^X(B_n) \qquad \square$$

Wesentlich in Def. 3.1: 'meßbar', d.h. ZV'en X induzieren auf $(\mathbb{R}^1, B^1)$ die Verteilung P^X.
Mathematisches Modell für Zufallsexperimente häufig: $(\Omega, \mathfrak{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}, P^X)$
Dabei modelliert die **linke Seite** alle Zufallseinflüsse, genaue Kenntnis oft nicht erforderlich, lediglich Existenz.
Die **rechte Seite** modelliert die interessanten beobachteten Größen. P^X oft bekannt bis auf Parameter. Beachte: $(\mathbb{R}, \mathcal{B}, P^X)$ ist ein W'Raum, bisherige Regeln gelten auch hier.

Mittels ZV X werden nur die uns interessierenden Dinge 'rausgefiltert', was $(\Omega, \mathfrak{A}, P)$ ist, ist uns egal, nur Ergebnisse sind von Relevanz. Dabei ist wichtig: Meßbarkeit!!!

Beispiel 3.3 (Binomialverteilung)
n-facher Münzwurf; Kopf$\widehat{=}$1, Zahl$\widehat{=}$0, Würfe unabhängig.
Kopf mit W'keit p, Zahl mit W'keit (1-p) in jedem Wurf.
Modell: $\Omega = \{\omega = (x_1, ..., x_n) | x_i \in \{0, 1\}\}$, $\mathfrak{A} = \mathcal{P}(\Omega)$,

$$P(\{(x_1, ..., x_n)\}) = \underbrace{p...p}_{|1'en|}\underbrace{(1-p)...(1-p)}_{|0'en|} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

ZV X = Anzahl der 1'en

$$X(\omega) = X((x_1, ...x_n)) = \sum_{i=1}^n x_i$$

Wertebereich von X: $T = \{0, 1, 2, ..., n\}$
Verteilung von X: $P^X(\{k\}) = P(X^{-1}(\{k\})) = P(\{\omega | X(\omega) = k\}) = P(X = k) =$

$$P(\{(x_1, ..., x_n)| \sum_{i=1}^n x_i = k\}) = \sum_{(x_1,...x_n): \sum_{i=1}^n x_i = k} p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

$$= \sum_{(x_1,...x_n): \sum_{i=1}^n x_i = k} p^k(1-p)^{n-k} = \binom{n}{k} p^k(1-p)^{n-k}, k = 0, 1, ..., n$$

Eine ZV X heißt **binomialverteilt** mit 2 Parametern n $\in \mathbb{N}, p \in [0, 1]$, wenn $P(X = k) = \binom{n}{k} p^k(1-p)^{n-k}$, $k = 0, 1, ..., n$
Bezeichnung: $Bin(n, p)$

Interpretation: $Bin(n,p)$ ist die Verteilung der Anzahl der 'Treffer' in Bernoulli-Serie[5] der Länge n mit Trefferwahrscheinlichkeit p.
Im Weiteren: Allgemeine Methoden zur Beschreibung von Wahrscheinlichkeitsverteilungen, direkt formuliert für ZV'en. Beachte: Mit X = Identität lassen sich die folgenden Überlegungen auch direkt auf W'Maße P anwenden.

3.1 Diskrete Verteilungen / Zufallsvariablen

Definition 3.4 Eine ZV X (auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$) bzw. deren Verteilung P^X heißt **diskret**, wenn eine **höchstens abzählbare Menge** T $= \{t_1, t_2, ...\}$ ex. mit $P^X(T) = 1 = P(X \in T)$.

X sei diskrete ZV mit Träger $T = \{t_1, t_2, ...\}, A \in \mathcal{B}^1$. Dann gilt $P^X(A \cap T) = P^X(A) + P^X(T) - P^X(A \cup T) = P^X(A)$
Also: $P(X \in A) = P^X(A) = P^X(A \cap T) = P^X(\bigcup_{i=1}^{\infty}(A \cap \{t_i\})) = \sum\limits_{i:t_i \in A} P^X(\{t_i\})$
$= \sum\limits_{i:t_i \in A} P(X = t_i)$

Definition 3.5 X sei diskrete ZV mit Träger $T = \{t_1, t_2, ...\}$. Die Abbildung $f_X : T \to [0,1]$ mit $f_X(t_i) = P(X = t_i)$, $i = 1, 2, ...$ heißt **Zähldichte** der ZV'en X.

$X \sim Bin(n,p)$ (s.Bsp. 3.3) ist ein Beispiel für eine diskrete ZV mit
$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, ..., n$, $(0 \leq p \leq 1 \text{ ein Parameter})$

Beispiel 3.6 (geometrische Verteilung)
Unendlicher, unabhängiger Münzwurf, s. Beispiel 3.3; $\Omega = \{\omega = (x_1, x_2, ...) | x_i \in \{0, 1\}\}$;
X = Wartezeit bis zum ersten Treffer ('1'); Träger $T = \mathbb{N}_0$. $P(X = k) = P(\{\omega = (x_1, x_2, ...) | x_1 = x_2 = ... = x_k = 0, x_{k+1} = 1\}) = (1-p)^k p$, $k = 0, 1, 2, ...$
Die Verteilung mit der Zähldichte $f_X(k) = (1-p)^k p$, $k = 0, 1, 2, ...$ und $(0 \leq p \leq 1$ *ein Parameter*) heißt **geometrische Verteilung**.
Bezeichnung: $X \sim Geo(p)$, $0 \leq p \leq 1$

Beispiel 3.7 (Poissonverteilung)
Sei $p_n \in (0,1)$ mit $np_n \to \lambda$ (für $n \to \infty$), $\lambda > 0$. Dann gilt $\binom{n}{k} p_n^k (1-p_n)^{n-k} \to e^{-\lambda} \frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N}_0$
Es gilt $e^{-\lambda} \frac{\lambda^k}{k!} \geq 0$ und $e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{e^{\lambda}} = 1$
Eine diskrete ZV'e X mit der Zähldichte $f_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, 2, ...$ heißt

[5] Als eine 'Bernoulli-Serie' bezeichnet man z.B. eine Münzwurfserie der Länge n mit Trefferwahrscheinlichkeit p, wobei die einzelnen Würfe unabhängig voneinander sind.

poissonverteilt ($\lambda > 0$ ein Parameter)
Bezeichnung: $X \sim Poi(\lambda)$, $\lambda > 0$ Interpretation: Man unterteile das Intervall $(0,1)$ in n Stücke (gleicher Länge). Die Wahrscheinlichkeit für das Auftreten eines Ereignisses in jedem Teilstück ist $\frac{\lambda}{n}$, das Auftreten ist unabhängig. Werden nun die Intervalle immer kürzer gewählt und die Wahrscheinlichkeit für die einzelen Ereignisse immer geringer, so wird die Poissonverteilung zur Beschreibung benutzt (Gesetzt seltener Ereignisse).

3.2 Verteilungsfunktionen

Definition 3.8 Eine Funktion $F : \mathbb{R} \to [0,1]$ mit den Eigenschaften

(i) F monoton steigend (nicht notwendig streng)
(ii) $\lim\limits_{x\to\infty} F(x) = 1,\ \lim\limits_{x\to\infty} F(x) = 0$
(iii) F ist rechtsseitig stetig (d.h. $\forall x_0$, $x_n \downarrow x_0$: $F(x_n) \to F(x_0)$)

heißt **Verteilungsfunktion** (VF).

Beispielskizzen siehe Mitschrift!

Satz 3.9 $(\Omega, \mathfrak{A}, P)$ sei W'Raum. $X : (\Omega, \mathfrak{A}) \to (\mathbb{R}^1, \mathcal{B}^1)$ sei ZV'e. Durch $F_X(x) = P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}) = P^X((-\infty, x])$, $x \in \mathbb{R}$ wird eine Verteilungsfunktion definiert, die **Verteilungsfunktion der ZV'en X**, bzw. der Verteilung P^X.

Satz 3.10 (Eindeutigkeitssatz für VF)
2 ZV'en X und Y besitzen dieselbe Verteilung (auf $(\mathbb{R}^1, \mathcal{B}^1)$) $\iff$ $F_X(x) = F_Y(x) \quad \forall x \in \mathbb{R}$

D.h. nicht, daß X=Y; punktweises Verhalten muß nicht gleich sein.

Beweis:
$'\Longrightarrow'$ einfach
$'\Longleftarrow'$ benutzt den Fortsetzungssatz und Eindeutigkeitssatz der Meßtheorie (sehr schwierig!) $\square$

Beispiel 3.11 (VF)

a) X heißt **gleichverteilt** (rechteckverteilt) auf [0,1] Bez. $X \sim R(0,1)$, wenn
$$F_X(x) = \begin{cases} 0 & : x \leq 0 \\ x & : 0 \leq x \leq 1 \\ 1 & : x \geq 1 \end{cases}$$

Seien $0 \leq a < b \leq 1$

$P(a < X \le b) = P(\{\omega | a < X(\omega) \le b\}) =$
$P(\{\omega | X(\omega) \le b\} \setminus \{\omega | X(\omega) \le a\}) =$
$P(\{\omega | X(\omega) \le b\}) - P(\{\omega | X(\omega) \le a\}) = P(X \le b) - P(X \le a) =$
$F_X(b) - F_X(a) = b - a$
(Wahrscheinlichkeit für zufälligen Wert im Intervall (a,b] ist gleich der Länge des Intervalls von (a,b], nämlich b-a)

Analog: Rechteckverteilung auf (a,b], Bez. $X \sim R(a,b) \quad a < b \in \mathbb{R}$

$$F_X(x) = \begin{cases} 0 & : x \le a \\ \frac{1}{b-a}(x-a) & : a \le x \le b \\ 1 & : x \ge b \end{cases}$$

Für ein Teilintervall (c,d] von (a,b] gilt:
$P(c < X \le d) = F_X(d) - F_X(c) = F_X(d) - F_X(c) = \frac{1}{b-a}(d-a) - \frac{1}{b-a}(c-a) = \frac{1}{b-a}(d-c)$
Wahrscheinlichkeit des Intervalls ist proportional zu deren Länge. Der Faktor beträgt $\frac{1}{b-a}$, was der Steigung der Funktion entspricht.

b) X heißt **exponentialverteilt**, $X \sim Exp(\lambda) \quad (\lambda > 0$ ein Parameter), wenn

$$F_X(x) = \left\{ \begin{array}{ll} 1 - e^{-\lambda x} & : x \ge 0 \\ 0 & : x \le 0 \end{array} \right\} = (1 - e^{-\lambda x}) 1_{[0,\infty)}(x)$$

Je größer λ, je steiler ist die Fkt., je kleiner das λ, umso flacher ist sie.

Bei dieser Darstellung wird benutzt:
Indikatorfunktion $\mathbf{1}_A(x) = \begin{cases} 1 & : x \in A \\ 0 & : x \notin A \end{cases}$

c) Diskrete ZV'en X mit geordnetem Träger T=$\{t_1, t_2, ...\} \subset \mathbb{R}$. Zähldichte
$f_X(t_i) = p_i \quad , p_i \ge 0 \quad , \sum_{i=1}^{\infty} p_i = 1.$
Dann $F_X(x) = P(X \le x) = \sum_{k: t_k \le x} P(X$
$= t_k) = \sum_{k: t_k \le x} p_k = \sum_{j=1}^{k-1} p_j$, falls $t_{k-1} \le x \le t_k$, $k = 1, 2, ...(t_0 = -\infty)$
Diese Fkt. stellt eine Treppenfunktion dar. Der Sprung erfolgt an der Stelle t_i und die Höhe des Sprungs entspricht dem Wert von p_i.

z.B.: $X \sim Geo(p)$, $f_X(k) = (1-p)^k p$, $k \in \mathbb{N} (0 < p \le 1)$
'Wartezeit bis zum ersten Treffer!'
Es gilt: $\sum_{j=0}^{k} (1-p)^j p = p \sum_{j=0}^{k} (1-p)^j = p \frac{1-(1-p)^{k-1}}{1-(1-p)} = 1 - (1-p)^{k+1}$

Also: $F_X = \begin{cases} 0 & : x < 0 \\ 1 - (1-p)^{\lfloor x \rfloor + 1} & : x \ge 0 \end{cases}$

Berechnung von Wahrscheinlichkeiten und Verteilungsfunktionen
X sei ZV'e mit Verteilungsfunktion $F_X(x) = P(X \le x)$, $a < b \in \mathbb{R}$

- $P(a < X \leq b) = P^X((-\infty, b] \setminus (-\infty, a]) \overset{Monotonie}{=} P^X((-\infty, b]) - P^X((-\infty, a]) = F_X(b) - F_X(a)$
- Sei $a \in \mathbb{R}$, $a_n < a_{n+1} < a$, $a_n \to a$ $(n \to \infty)$ Dann: $(a_n, a]$ ist absteigend mit $\lim\limits_{n\to\infty}(a_n, a] = \{a\}$
  $P(X = a) = P^X(\{a\}) = P(\lim\limits_{n\to\infty}(a_n, a]) = \lim\limits_{n\to\infty} P((a_n, a]) = \lim\limits_{n\to\infty}(F_X(a) - F_X(a_n)) = F_X(a) - \lim\limits_{n\to\infty} F_X(a_n) = F_X(a) - \underbrace{F_X(a-)}_{\text{linksseitiger Grenzwert von } F_X(a)}$
  Insbesondere: Ist F_X stetig, so gilt $P(X = a) = 0 \quad \forall a \in \mathbb{R}$
- $P(a \leq X \leq b) = P(X = a) + P(a < X \leq b) = F_X(a) - F_X(a-) + F_X(b) - F_X(a) = F_X(b) - F_X(a-)$

Bedienzeit X von Anforderungen an einem Server sei $Exp(\lambda)$-verteilt
$X \sim Exp(\lambda) \quad \lambda \geq 0$
Problem: Bestimme die Zeit x_α, unterhalb derer die Bedienzeit mit vorgegebener Wahrscheinlichkeit α liegt. (Z.B. $\alpha = 0,99$)
Bestimme also x_α mit $P(X \leq x_\alpha) = \alpha$.
Interpretation: In 99% der Fälle ist die Bedienzeit kürzer als x_α

Definition 3.12 X sei ZV'e mit Verteilungsfunktion $F_X(x)$, $0 \leq \alpha \leq 1$. $x_\alpha = \min\{x | F_X(x) \geq \alpha\}$ heißt α**-Quantil** ((1-α)-Fraktil) von F_X.
$F_X^-(t) = \min\{x | F_X(x) \geq t\}$, $t \in (0,1)$ heißt **Pseudoinverse** von F_X.

Ist F invertierbar, so gilt $F^- = F^-$ (Inverse)
Insbesondere: $\alpha = \frac{1}{2}$, $x_{\frac{1}{2}}$ heißt **Median**[6] von F_X. Es gilt: $P(X \leq x_{\frac{1}{2}}) \leq \frac{1}{2}$. Der Median 'halbiert' die Verteilung, in diesem Sinn der 'mittlere' Wert.

3.3 Dichten

Methode zur Konstruktion von Verteilungsfunktionen, Verteilungen.

Definition 3.13 $f : \mathbb{R} \to \mathbb{R}^+$ sei (uneigentlich Riemann-) integrierbar, mit $\int\limits_{-\infty}^{\infty} f(x)dx = 1$. Durch $F(x) = \int\limits_{-\infty}^{x} f(t)dt$ wird eine Verteilungsfuntion definiert.
Gilt für eine ZV X : $F_X(x) = F(x)$, so heißt f **(Verteilungs-)Dichte** von X bzw. P^X. X (bzw. P^X) heißt dann **alsolut-stetig**.

Mit dem Hauptsatz der Differential- und Integralrechnung (HDI) folgt:
$f(x) = F'(x)$ für alle Stetigkeitspunkte x von f.

[6] Median ist ein Schätzer für den mittleren Wert, der robust gegen Ausreißer ist.

Beispiel 3.14

a) Rechteckverteilung auf $[a,b]$, $a < b \in \mathbb{R}$, $X \sim R(a,b)$:
$f(x) = \begin{cases} \frac{1}{b-a} & : a \leq x \leq b \\ 0 & : sonst \end{cases} = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x),\ x \in \mathbb{R}$
$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \begin{cases} \frac{1}{b-a}(x-a) & : a \leq x \leq b \\ 0 & : x \leq a \\ 1 & : x \geq b \end{cases}$
Beachte: Die Dichte $\hat{f}(x) = \frac{1}{b-a}\mathbb{1}_{(a,b)}(x)$ führt zu derselben Verteilungsfunktion F. Dichten sind nicht eindeutig, sie sind 'fast sicher' eindeutig.

b) Exponentialverteilung, $Exp(\lambda)$, $X \sim Exp(\lambda)$, $\lambda \geq 0$
$f(x) = \lambda e^{-\lambda x}\mathbb{1}_{[0,\infty)},\ x \in \mathbb{R}$
$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{0}^{x} \lambda e^{-\lambda t}dt$ (Da Funktion links nicht definiert!)
$= -e^{-\lambda t}|_0^x = 1 - e^{-\lambda x},\ x \geq 0$

c) Normalverteilung, $N(\mu,\sigma^2)$, $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+\backslash\{0\}$, $X \sim N(\mu,\sigma^2)$
$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in \mathbb{R}$
$F(x) = \frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$ hat keine 'geschlossene' Darstellung, allerdings interessiert man sich für explizite Werte der Funktion. Berechnung dieser erfolgt entweder numerisch oder mit Hilfe von Tabellen.

Berechnung von Wahrscheinlichkeiten mit Dichten.
X sei ZV'e mit Verteilungsfunktion F_X und Dichte $f_X(x)$, $a < b \in \mathbb{R}$

- $P(a < X \leq b) = F_X(b) - F_X(a) = \int\limits_{-\infty}^{b} f_X(t)dt - \int\limits_{-\infty}^{a} f_X(t)dt = \int\limits_{a}^{b} f_X(t)dt$

- $P(X = a) = F_X(a) - F_X(a^-) = 0$, da $F_X(a) = \int\limits_{-\infty}^{x} f_X(t)dt$ stetig

- $P(a \leq X \leq b) = \int\limits_{a}^{b} f_X(t)dt$, da einzelne Punkte die W'keit 0 haben.

- Allgemein gilt bei absolut-stetigen ZV'en:
  $P \in (X \in < a,b >) = \int\limits_{a}^{b} f_X(t)dt$, wobei $<>$ für 'abgeschlossen' oder 'offen' steht.

- Schreibe allgemein für Mengen $B \in \mathcal{B}$:
  $P(X \in B) = \int\limits_{B} f_X(t)dt$ (auch, wenn B kein Intervall)

3.4 Erzeugende Funktionen und LaPlace-Transformierte

Eine weitere Methode zur eindeutigen Beschreibung von Verteilungen.

Definition 3.15

a) X sei diskrete ZV'e mit Träger $T = \{t_1, t_2, ...\}$ und Zähldichte $f_X(t_k) = p_k$, $k \in \mathbb{N}$. Dann heißt: $G_X(z) = \sum\limits_{k=0}^{\infty} p_k z^k$, $|z| < 1$, **erzeugende Funktion** von X bzw. P^X

b) X sei absolut-stetig mit Dichte f_X, wobei $f_X(x) = 0$, falls $x < 0$ (d.h. $P(X < 0) = 0$). Dann heißt: $L_X(s) = \int\limits_0^{\infty} e^{-sx} f_X(x)dx$, $s \geq 0$ **LaPlace-Transformierte** von X bzw. P^X.

Analog zu Satz 3.10 für Verteilungsfunktionen gilt auch hier Eindeutigkeit.

Satz 3.16

a) X, Y seinen diskrete ZV'en mit demselben Träger T. X und Y besitzen dieselbe Verteilung $\Longleftrightarrow G_X(z) = G_Y(z)$, $\forall |z| \leq 1$
b) X, Y seien absolut-stetig mit $f_X(x) = f_Y(x) = 0$, $\forall x < 0$. X und Y haben dieselbe Verteilung $\Longleftrightarrow L_X(s) = L_Y(s)$, $\forall s \geq 0$

Beweis:

a) Eindeutigkeitssatz für Potenzreihen
b) Felber II, Seite 403

Satz 3.17 (Inversionsformel)

a) $G_X(z)$ sei erzeugende Funktion einer diskreten ZV'en X mit Träger $T = \{t_1, t_2, ...\}$. Dann gilt: $P(X = t_k) = \frac{1}{k!} G_X^{(k)}(0)$.
b) $L_X(s)$ sei LaPlace-Transformierte einer absolut stetigen ZV'en X. Dann gilt: $f_X(x) = \lim\limits_{y \to \infty} \frac{1}{2\pi i} \int\limits_{c-iy}^{e+iy} e^{sx} L_X(s) ds$, $\forall c \in \mathbb{R}$

Beispiel 3.18

a) Geometrische Verteilung, $X \sim Geo(p)$, $0 < p \leq 1$
$G_X(z) = \sum\limits_{k=0}^{\infty} (1-p)^k p z^k = p \sum\limits_{k=0}^{\infty} ((1-p)z)^k = \frac{p}{1-z+pz}$, $|z| < 1$
b) Poissonverteilung, $X \sim Poi(\lambda)$, $\lambda > 0$
$G_X(z) = \sum\limits_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} z^k = e^{-\lambda} \sum\limits_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = e^{-\lambda} e^{\lambda z} = e^{-\lambda(1-z)}$, $z \in \mathbb{R}$

c) Rechteckverteilung, $X \sim R(0,1)$

$L_X(s) = \int\limits_0^1 e^{-sx} dx = \frac{1-e^{-s}}{s},\ s \geq 0$

d) Exponentialverteilung, $X \sim Exp(\lambda)$

$L_X(s) = \int\limits_0^\infty e^{-sx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda+s} \underbrace{\int\limits_0^\infty (\lambda+s) e^{-(s+\lambda)x} dx}_{\text{1, da Integral über Dichte}} = \frac{\lambda}{\lambda+s}$

4 Produkträume und Zufallsvektoren

Zufallsexperimente mit mehreren Ausgängen (gleichzeitig oder dasselbe Experiment mehrfach wiederholt)

$$\begin{tikzcd}
 & (\mathbb{R}^1, \mathcal{B}^1) \\
(\Omega, \mathfrak{A}, P) \arrow[ur] \arrow[r] \arrow[dr] & (\mathbb{R}^1, \mathcal{B}^1) \\
 & \vdots \\
 & (\mathbb{R}^1, \mathcal{B}^1)
\end{tikzcd}$$

Modell: Ausgänge gemeinsam betrachten:

$(\Omega, \mathfrak{A}, P) \xrightarrow{X=(X_1,...X_n)} (\mathbb{R}^n, ?, ?)$

<u>Fragen</u>

1. Welche σ-Algebra?
2. Welche gemeinsame Verteilung?
3. Wie beschreiben?

4.1 Produkträume

Gegeben sind Meßräume $(\Omega_i, \mathfrak{A}_i)$, $i = 1, ..., n$. Oft $(\Omega_i, \mathfrak{A}_i) = (\mathbb{R}^1, \mathcal{B}^1)$.
Setze $\Omega = \bigtimes_{i=1}^{\infty} \Omega_i = \{w = (w_1, ..., w_n) | w_i \in \Omega_i\}$
Frage: Welche σ-Algebra ist über Ω zu wählen?
Ansatz: $\varepsilon = \{A_1 \times ... \times A_n | A_i \in \Omega_i,\ i = 1, ..., n\}$ Aber: ε ist i.a. keine σ-Algebra über $\Omega = \bigtimes_{i=1}^{n} \Omega_i$
Gegenbeispiel: $\Omega_1 = \Omega_2 = \mathbb{R}^1$, $\mathfrak{A}_1 = \mathfrak{A}_2 = \mathcal{B}^1$. Dann: $C_1 = [-1, 0] \times [-1, 0] \in \varepsilon$, $C_2 = [0, 1] \times [0, 1] \in \varepsilon$; aber $C_1 \cup C_2 \notin \varepsilon$, nicht als karthesisches Produkt zweier Mengen darstellbar.

Definition 4.1 (Produkt-σ-Algebra)
$(\Omega_i, \mathfrak{A}_i)$ seien Meßräume, $i = 1, ..., n$; $\varepsilon = \{A_1 \times ... \times A_n | A_i \in \mathfrak{A}_i\}$. $\bigotimes_{i=1}^n \mathfrak{A}_i := \mathfrak{A}(\varepsilon)$ heißt **Produkt-σ-Algebra** von $\mathfrak{A}_1, ..., \mathfrak{A}_n$ über $\Omega = \bigtimes_{i=1}^{n} \Omega_i$. $(\bigtimes_{i=1}^{n} \Omega_i, \bigotimes_{i=1}^n \mathfrak{A}_i)$ heißt **Produkt-Meßraum der** $(\Omega_i, \mathfrak{A}_i)$**,** $i = 1, ..., n$

Beispiel 4.2 $(\Omega_i, \mathfrak{A}_i) = (\mathbb{R}^1, \mathcal{B}^1)$, $\varepsilon = \{B_1 \times ... \times B_n | B_i \in \mathcal{B}^1\}$. $\mathfrak{A}(\varepsilon) = \mathcal{B}^n$ heißt dann **n-dimensionale Borelsche σ-Algebra**.
Z.B.:

- enthält $\mathcal{B}^n$ alle (n-dim) Rechtecke oder
- alle abzählbaren Vereinigungen von Rechtecken oder
- alle Linien (durch abz. Vereinigung oder Schnitt von Rechtecken) oder

- alle Kreise (um den Ursprung): $\{x_1^2 + x_2^2 \leq c\}$ (hier $\mathbb{R}^2$)

4.2 Zufallsvektoren und Folgen von Zufallsvariablen

Definition 4.3 $(\Omega, \mathfrak{A}, P)$ sei Wahrscheinlichkeitsraum. Eine Abbildung $X = (X_1, ..., X_n)$, $\Omega \to \mathbb{R}^n$ heißt **Zufallsvektor**, wenn $X^{-1}(B) \in \mathfrak{A} \ \ \forall B \in \mathcal{B}^n$ (Meßbarkeit).
Bez.: $X : (\Omega, \mathfrak{A}) \to (\mathbb{R}^n, \mathcal{B}^n)$

Bemerkung: Es gilt: $X = (X_1, ..., X_n)$ ist Zufallsvektor $\Longleftrightarrow X_1, ..., X_n$ sind ZV'en.

Lemma 4.4 X sei Z'vektor. Durch $P^X(B) = P(X^{-1}(B)) = P(\{\omega | X(\omega) \in B\}) \overset{kurz}{=} P(X \in B)$, $B \in \mathcal{B}^n$ wird eine Wahrscheinlichkeitsverteilung auf $(\mathbb{R}^n, \mathcal{B}^n)$ definiert. $P^X = P^{(X_1,...,X_n)}$ heißt **gemeinsame Verteilung** von $(X_1, ..., X_n)$

Im Folgenden gemeinsame Verteilung mit Verteilungsfunktion und Dichten beschreiben. Klar: $X_1, ..., X_n$ diskret $\Longrightarrow (X_1, ..., X_n)$ diskret. Verteilung mit Zähldichte wie in Kapitel 3 beschreiben.

Satz 4.5 X sei Zufallsvektor, P^X wird eindeutig beschrieben durch die (n-dimensionale) Verteilungsfunktion: $F_X(x_1, ..., x_n) = P^X((-\infty, x_1] \times ... \times (-\infty, x_n])$
$= P(\{\omega | X_1(\omega) \in (-\infty, x_1], ..., X_n(\omega) \in (-\infty, x_n]\})$
$= P(\{\omega | X_1(\omega) \leq x_1\} \cup ... \cup \{X_n(\omega) \leq x_n\}) \overset{kurz}{=} P(X_1 \leq x_1, ..., X_n \leq x_n)$

Bez.: $X \sim F_X$

Beschreibung von Verteilungen mit n-dimensionalen Dichten ist oft einfacher.

Definition 4.6 $F_X(x_1, ..., x_n)$ sei die Verteilungsfunktion des Zufallsvektors X. Eine (uneigentlich Riemann-) integrierbare Funktion $f_X : \mathbb{R}^n \to \mathbb{R}^+$ heißt **Dichte von X**, wenn $F_X(x_1, ..., x_n)) = \int\limits_{-\infty}^{x_n} \cdots \int\limits_{-\infty}^{x_1} f_X(t_1, ..., t_n) dt_1 \cdots dt_n$, $\forall x_1, ...x_n \in \mathbb{R}$.
X (bzw. F_X oder P^X) heißt dann absolut-stetig mit Dichte f_X.

Bez.: $X \sim f_X$
Sprechweisen: X hat/besitzt/ist verteilt nach Verteilungsfunktion F_X/ Dichte f_X.

Berechnung von Wahrscheinlichkeiten mit Dichten:
$X = (X_1, ..., X_n)$ sei absolut-stetig mit Verteilung P^X, Verteilungsfunktion F_X und Dichte f_X.
Dann: $P^X(< a_1, b_1 > \times ... \times < a_n, b_n >) = P(a_1 \leq X_1 \leq b_1, ..., a_n \leq X_n \leq b_n) = \int\limits_{a_n}^{b_n} \cdots \int\limits_{a_1}^{b_1} f_X(t_1, ..., t_n) dt_1 \cdots dt_n$, $\forall a_i \leq b_i \in \mathbb{R}$, wobei '<' = '(' oder '[' ('>' analog)

beliebig auswählbar.
Allg.: $P^X(B) = \int\underset{B}{\cdots}\int f_X(t_1, ..., t_n)dt_1 \cdots dt_n, \forall B \in \mathcal{B}^n$
Mit dem HDI: $f_X(x_1, ..., x_n) = \frac{\partial}{\partial x_1 \cdots \partial x_n} F_X(x_1, ..., x_n)$ in allen Stetigkeitspunkten $(x_1, ...x_n)$ von F.

Beispiel 4.7 (Indizierung mit X zur Vereinfachung weggelassen)

a) Gleichverteilung auf $T \in \mathcal{B}^n$ mit $c = \int\underset{T}{\cdots}\int 1dx_1 \cdots dx_n \leq \infty$

$f(x_1, ..., x_n) = \frac{1}{c}\mathbb{1}_T(x_1, ..., x_n) = \begin{cases} \frac{1}{c}: & (x_1, ..., x_n) \in T \\ 0: & sonst \end{cases}$

Speziell:

- auf dem Einheitswürfel des $\mathbb{R}^n$: $T = \{(x_1, ..., x_n) | 0 \leq x_i \leq 1,\ i = 1, ..., n\} = [0, 1]^n$
  $f(x_1, ...x_n) = \mathbb{1}_T(x_1, ...x_n)$
  $F(x_1, ...x_n) = \begin{cases} 0: & \exists i : x_i < 0 \\ min\{x_1, 1\} \cdot ... \cdot min\{x_1, 1\}: & sonst \end{cases}$
- auf der Einheitskugel des $\mathbb{R}^n$
  $T = \{(x_1, ..., x_n) | \sum_{i=1}^{n} x_i^2 \leq 1\}$, $c = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}$ (Volumen der Einheitskugel)
  $\boxed{\begin{array}{l} \Gamma(n) = (n-1)! \quad (n \in \mathbb{N}) \\ \Gamma(\frac{1}{2}) = \sqrt{\pi} \qquad \Gamma(x+1) = x\Gamma(x) \end{array}}$
  $f(x_1, ..., x_n) = \frac{\Gamma(\frac{n}{2}+1)}{\pi^{\frac{n}{2}}}\mathbb{1}_T(x_1, ...x_n),\ (x_1, ...x_n) \in \mathbb{R}^n$
  Verteilungsfunktion ist schwierig zu berechnen → Dirichlet-Integrale

b) n-dim. Normalverteilung $\mathcal{N}(\mu, \Sigma)$, $\mu \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n\times n}$ positiv definit, d.h. EW> 0
$f(x_1, ..., x_n) = \frac{1}{(2\Pi)^{\frac{n}{2}}(det\Sigma)^{\frac{1}{2}}} exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$
$x = (x_1, ..., x_n)^T \in \mathbb{R}^n$

c) $f_1, ..., f_k : \mathbb{R}^n \to \mathbb{R}^+$ seien n-dim. Dichten (auch $n = 1$), $\alpha_1, ..., \alpha_k \geq 0$. $\sum_{i=1}^{k} \alpha_i = 1$. Dann ist $f(x) = \sum_{i=1}^{k} \alpha_i f_i(x)$, $x = (x_1, ..., x_n) \in \mathbb{R}^n$ eine n-dim. Dichte. $f(x)$ heißt **Mischung** der Dichten $f_1, ..., f_k$

Stochastische Unabhängigkeit

Definition 4.8 $X_1, ..., X_n$ seien ZV'en. Dann ist $X = (X_1, ..., X_n)$ ein Z'Vektor. $X_1, ..., X_n$ heißen **stochastisch unabhängig**, wenn $F_X(x_1, ..., x_n) = F_{X_1}(x_1) \cdot ... \cdot F_{X_n}(x_n)$, $\forall (x_1, ..., x_n) \in \mathbb{R}^n$

Also: $X_1, ..., x_n$ st. unabh., wenn $P(X_1 \leq x_1, ..., X_n \leq x_n) = P(X_1 \leq x_1) \cdot ... \cdot P(X_n \leq x_n)$, $\forall (x_1, ..., x_n) \in \mathbb{R}^n$

Oder:
$P(\{\omega|X_1(\omega) \leq x_1\} \cap ... \cap \{\omega|X_n(\omega) \leq x_n\}) = P(\{\omega|X_1(\omega) \leq x_1\} \cdot ... \cdot P(\{\omega|X_n(\omega) \leq x_n\})$ dh., $X_1, ...X_n$ sind st. unabh. $\Longleftrightarrow$ die Ereignisse $\{\omega|X_i(\omega) \leq x_i\}$, $i = 1, ..., n$, $\forall x_1, ..., x_n$ st. unabh. im Sinne der Def. 2.17 sind.

Lemma 4.9 $(X_1, ..., X_n)$ sei Z'Vektor, absolut stetig mit Dichte $f_{(X_1,...,X_n)}$. Dann hat jedes der X_i eine Dichte
$f_{X_i}(x) = \int\limits_{-\infty}^{\infty} ... \int\limits_{-\infty}^{\infty} f_{(X_1,...,X_n)}(x_1, ..., x_{i-1}, x, x_{i+1}, ..., x_n)dt_{x_1}...dt_{x_{i-1}}dt_{x_{i+1}}...dt_{x_n}$

Es gilt: $f_{(X_1,...,X_n)}(x_1, ..., x_n) = \prod\limits_{i=1}^{n} f_{X_i}(x_i)$, $\forall x_1, ..., x_n \in \mathbb{R} \Longrightarrow X_1, ..., X_n$ stochastisch unabhängig.
Umgekehrt: Sind $X_1, ..., X_n$ st. unabh. und absolut stetig mit Dichten f_{X_i}, so ist $f_{(X_1,...,X_n)}(x_1, ..., x_n) = \prod\limits_{i=1}^{n} f_{X_i}(x_i)$, $x_1, ..., x_n \in \mathbb{R}$ eine Dichte des Z'Vektors $(X_1, ..., X_n)$.

Beweis: $f_{X_i}(x) = \int\limits_{-\infty}^{\infty} ... \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} ... \int\limits_{-\infty}^{\infty} f_{(X_1,...,X_n)}(x_1, ..., x_{i-1}, x, x_{i+1}, ..., x_n)dt_{x_1}...$ $dt_{x_{i-1}}dt_{x_{i+1}}...dt_{x_n}$ ist eine Dichte von X_i, da $F_{X_i}(x) = P(X_i \leq x) = \int\limits_{-\infty}^{x} f_{X_i}(t_i)dt_i$
Es gilt: $F_{(X_1,...,X_n)}(x_1, ..., x_n) = \int\limits_{-\infty}^{x_n} ... \int\limits_{-\infty}^{x_1} \prod\limits_{i=1}^{n} f_{X_i} dt_1...dt_n = \prod\limits_{i=1}^{n} \int\limits_{-\infty}^{x_i} f_{X_i}(t_i)dt_i) = \prod\limits_{i=1}^{n} F_{X_i}(x_i)$, $\forall x_1, ..., x_n \in \mathbb{R}$
Also: $X_1, ..., X_n$ st. unabh. (Def. 4.8)
Umgekehrt: $X = (X_1, ..., X_n)$ st. unabh. $\Longrightarrow F_{(X_1,...,X_n)}(x_1, ..., x_n) = \prod\limits_{i=1}^{n} F_{X_i}(x_i) = \prod\limits_{i=1}^{n} \int\limits_{-\infty}^{x_i} f_{X_i}(t_i)dt_i = \int\limits_{-\infty}^{x_n} ... \int\limits_{-\infty}^{x_1} \prod\limits_{i=1}^{n} f_{X_i}(t_i)dt_i \Longrightarrow \prod\limits_{i=1}^{n} f_{X_i}(t_i)$ ist eine Dichte von X $\square$

Im Fall von Zähldichten vereinfacht sich Lemma 4.9 zu:

Lemma 4.10 $X_1, .., X_n$ seien diskrete ZV'en mit Trägern $T_1, ..., T_n$. Dann gilt: $X_1, .., X_n$ st. unabh. $\Longleftrightarrow P(X_1 = t_1, ..., X_n = t_n) = \prod\limits_{i=1}^{n} P(X_i = t_i)$, $\forall t_i \in T_i$

Definition 4.11 Eine Folge von ZV'en $\{X_n\}_{n \in \mathbb{N}}$ auf W'Raum $(\Omega, \mathfrak{A}, P)$ heißt **stochastisch unabhängig**, wenn $X_1, ..., X_n$ stoch. unabh. sind $\forall n \in \mathbb{N}$. Besitzen stoch. unabh. ZV'en alle dieselbe Verteilung, so heißen sie **stoch. unabh. identisch verteilt**. Kurz: *stid*

Beispiel 4.12

a) X_1, X_2 *stid* $\sim Bin(n,p)$.
$P(X_1 = X_2) = P^{(X_1,X_2)}(\{(i,i)|0 \le i,j \le n, i = j\}) = \sum\limits_{i=0}^{n} P^{(X_1,X_2)}(\{(i,j)\}) =$
$\sum\limits_{i=0}^{n} P(X_1 = i, X_2 = i) = \sum\limits_{i=0}^{n} P(X_1 = i)P(X_2 = i) = \sum\limits_{i=0}^{n}[\binom{n}{i}p^i(1-p)^{n-i}]^2$

b) X_1, X_2 stoch. unabh. abs.-stetig mit Dichten f_1, f_2
$P(X_1 = X_2) = \int\int\limits_{x_1=x_2} f(x_1,x_2)dx_1dx_2 = \int\limits_{-\infty}^{\infty}\int\limits_{x_2}^{x_2} f_1(x_1)dx_1f_2(x_2)dx_2 = 0$

Beispiel 4.13 $X_1, ..., X_n$ *stid* $\sim$ F (Verteilungsfunktion)

a) $Y = \max\{X_1, ..., X_n\}$ (Interpretation z.B. Max. der Laufzeiten)
$P(Y \le x) = P(X_1 \le x, ..., X_n \le x) = \prod\limits_{i=1}^{n} P(X_i \le x) = F^n(x)$, $x \in \mathbb{R}$.
Also $Y = \max\{X_1, ..., X_n\}$ hat Verteilungsfunktion $F^n(x)$.

b) $Y = min\{X_1, ..., X_n\}$
$P(Y > x) = P(X_1 > x, ..., X_n > x) = (1 - F(x))^n$
Also: $Y = min\{X_1, ..., X_n\}$ hat VF $1 - (1 - F(x))^n$
z.B.: X_i *stid* $\sim Exp(\lambda)$: $P(Y \le x) = 1 - e^{-n\lambda x}$, $\lambda > 0$; d.h.: $Y \sim Exp(n\lambda)$

Beispiel 4.14 X_1, X_2 *stid* $Exp(\lambda)$, $\lambda > 0$
$P(X_1 + X_2 \le z) = \int\int\limits_{0 \le x_1+x_2 \le z} \lambda e^{-\lambda x_1}\lambda e^{-\lambda x_2}\mathbb{1}_{[0,\infty)}(x_1)\mathbb{1}_{[0,\infty)}(x_2)dx_1dx_2$
$= \int\limits_0^z\int\limits_0^{z-x_2} \lambda e^{-\lambda x_1}\lambda e^{-\lambda x_2}dx_1dx_2 = \int\limits_0^z \lambda e^{-\lambda x_2}[1-e^{-\lambda(z-x_2)}]dx_2 = \int\limits_0^z(\lambda e^{-\lambda x_2}-\lambda e^{-\lambda z})dx_2$
$= 1 - e^{-\lambda z} - \lambda z e^{-\lambda z} = F_{X_1+X_2}(z)$, $z \ge 0$
Durch Differenzieren: $f_{X_1+X_2}(z) = \lambda e^{-\lambda z} - \lambda(e^{-\lambda z} + \lambda z e^{-\lambda z})$, $z \ge 0$
Also: Die Verteilung der Summe von stoch. unabh. identisch $Exp(\lambda)$-verteilten ZV'en ist absolut steitig verteilt mit Dichte $f(x) = \lambda^2 x e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x)$

Eine allgemeine Klasse von Verteilungen, die diese Dichte als Spezialfall enthält, sind die **Γ-Verteilungen** mit Dichten $f_{\alpha,\lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x)$, $\alpha, \lambda > 0$, wobei $\Gamma(\alpha) = \int\limits_0^\infty x^{\alpha-1}e^{-x}dx$ das **Γ-Integral** (Insbes. $\Gamma(n) = (n-1)!\ \forall n \in \mathbb{N}$). Eine ZV mit Dichte $f_{\alpha,\lambda}(x)$ heißt **$\Gamma(\alpha, \lambda)$-verteilt**.
Spezialfälle:

(i) $\alpha = 1 \to Exp(\lambda)$-Verteilung

(ii) $\alpha = 1 \to$ s.o. Verteilung der Summe von 2 *stid*, $Exp(\lambda)$-verteilten ZV'en

(iii) $\alpha = n \in \mathbb{N} \to$ Erlang-Verteilung ,mit Parameter n, λ
$f_{n,\lambda}(x) = \frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x)$

Beispiel 4.15 $\{X_n\}_{n\in\mathbb{N}}$ Folge von *stid* ZV'en, $B \in \mathcal{B}^1$
$S = min\{n \in \mathbb{N} | X_n \in B\}$ heißt erste **Eintrittszeit in** B. Verteilung von S:
$P(S = k) = P(X_1 \notin B, ..., X_{k-1} \notin B, X_k \in B) = P(X_1 \notin B) \cdot ... \cdot P(X_{k-1} \notin B) \cdot P(X_k \in B) = (1-p)^{k-1}p$, $k \in \mathbb{N}$, wobei $p = P(X_1 \in B) = ... = P(X_n \in B)$.
Also: S geometrisch verteilt mit Träger $\mathbb{N}$: $S - 1 \sim Geo(p)$
Konkret: Unabh. Folge von Münzwürfen,
X_n *stid* $\sim Bin(1,p)$, $B = \{1\}$, $P(X_n \in \{1\}) = P(X_n = 1) = p$
$\{S = k\}$: Beim k-ten Wurf tritt erstmalig 1 ('Treffer') auf. Setze $X = S - 1 =$ Wartezeit bis zum ersten Treffer. Dann: $X \sim Geo(p)$.

Satz 4.16 $X_1, ..., X_n$ seien stoch. unabh. ZV'en und $I, J \subseteq \{1, ..., n\}$, $I \cap J = \emptyset$. Dann sind $(X_i)_{i\in I}$ und $(X_j)_{j\in J}$ stoch. unabh. Z'Vektoren. Sind f, g meßbare Abbildungen (mit entsprechenden Bild-Meßräumen), so sind $f((X_i)_i \in I)$ und $f((X_j)_j \in J)$ stoch. unabh.

Beweis: Mathar und Pfeiffer (Lemma 2.1.6 und 2.1.7, Seite 74)

Bsp.: X_1, X_2, X_3, X_4 stoch. unabh. $\Longrightarrow X_1 + X_2$, $X_3 + X_4$ stoch. unabh.

5 Transformation von ZV'en/Verteilungen

Im folgenden: allg. Hilfsmittel zur Berechnung der Verteilung von Fkt. von ZV'en.

Satz 5.1 (Transformationssatz für Dichten)
$X = (X_1, ..., X_n)$ sei Z'Vektor auf W'Raum $(\Omega, \mathfrak{A}, P)$ absolut-stetig mit Dichte f_X. Gelte $f_X(x_1, .., x_n) = 0\ \forall (x_1, ..., x_n) \in M^C$ für eine offene Menge $M \subseteq \mathbb{R}^n$. $T : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^n, \mathcal{B}^n)$ sei eine meßbare Abbildung (Z'Vektor, $T^{-1}(B) \in \mathcal{B}^n\ \forall B \in \mathcal{B}^n$) mit

(i) $\tilde{T} = T|_M$ ist injektiv (Restriktion von T auf M)

(ii) $\tilde{T}$ ist stetig differenzierbar auf M

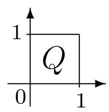
(iii) $det(\left\{ \frac{\partial \tilde{T}_i}{\partial x_j} \right\}_{1 \le i,j \le n}) \neq 0$ auf M (**Funktionaldeterminante**)
Dann ist $Y = T(X)$ absolut stetig mit Dichte

$$f_Y(y_1, ..., y_n) = \frac{f_X(\tilde{T}^{-1}(y_1, ..., y_n)) \mathbb{1}_{\tilde{T}(M)}(y_1, ..., y_n))}{\left| det(\left\{ \frac{\partial \tilde{T}_i}{\partial x_j} \right\}_{1 \le i,j \le n} \Big|_{\tilde{T}^{-1}(y_1,...,y_n)}) \right|}$$
$$= \left| det \left\{ \left(\frac{\partial \tilde{T}_i^{-1}}{\partial y_j} \right)_{1 \le i,j \le n} \right\} f_X(\tilde{T}^{-1}(y_1, ..., y_n)) \mathbb{1}_{\tilde{T}(M)}(y_1, ..., y_n) \right|$$

Absolutbetrag der Funktionaldeterminante an der Stelle $T^{-1}(y_1, ..., y_n)$

Erläuterung
$T_1(y_1, x_2, x_3) = x_1 x_2 x_3$
$T_1(y_1, x_2, x_3) = x_1 + x_2$ Dann z.B.: $\frac{\partial T_1}{\partial x_3} = x_1 x_2$
$T_1(y_1, x_2, x_3) = x_3$

Beispiel 5.2 $X = (X_1, X_2)$ gleichverteilt auf $Q = (0, 1)^2$ mit Dichte $f_X(x_1, x_2) = \mathbb{1}_Q(x_1, x_2)$, d.h. X_1, X_2 stoch. unabh. $X_i \sim R(0, 1)$, $T(x_1, x_2) = (\sqrt{x_1} cos(2\pi x_2), \sqrt{x_1} sin(2\pi x_2))$, $(x_1, x_2) \in Q$.

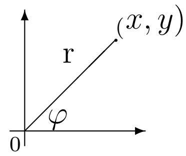


$$det \left\{ \frac{\partial T_i}{\partial x_j} \right\}_{1 \le i,j \le 2} = det \left\{ \begin{array}{cc} \frac{1}{2\sqrt{x_1}} \cos(2\pi x_2) & -2\pi\sqrt{x_1} \sin(2\pi x_2) \\ \frac{1}{2\sqrt{x_1}} \sin(2\pi x_2) & 2\pi\sqrt{x_1} \cos(2\pi x_2) \end{array} \right\}$$
$$= \pi \cos^2(2\pi x_2) + \pi \sin^2(2\pi x_2) = \pi(\cos^2(2\pi x_2) + \sin^2(2\pi x_2)) = \pi$$

$Y = T(X)$ besitzt nach Satz 5.1 eine Dichte $f_Y(y_1, y_2) = \frac{1}{\pi} f_X(T^{-1}(y_1, y_2)) \mathbb{1}_{T(Q)}(y_1, y_2)$ $= \frac{1}{\pi}(\mathbb{1}_{T(Q)}(y_1, y_2))^2 = \frac{1}{\pi} \mathbb{1}_{T(Q)}(y_1, y_2)$ mit $\hat{K} = T(Q) \mathrel{\hat{=}}$ Einheitskreis ohne positive x-Achse. Also $Y = T(X)$ ist gleichverteilt auf Einheitskreis. (vgl. Bsp. 4.7)

Beispiel 5.3 (Rayleigh-Verteilung)
$X, Y\, stid \sim N(0, \sigma^2)$, also gemeinsame Dichte $f_{(X,Y)}(x,y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}}$, $(x,y) \in \mathbb{R}^2$
Transformation auf Polarkoordinaten:
$(r, \Phi) = T(x,y) : \mathbb{R}^2 \backslash \{0\} \to (0,\infty) \times (0, 2\Phi)$ mit $r = \sqrt{x^2+y^2}$ (Länge)

$$\Phi = \begin{cases} \arctan \frac{y}{x} & y > 0 \\ \pi + \arctan \frac{y}{x} & y < 0 \end{cases} \quad \text{(Winkel) (mit } \arctan(^+_-\infty) = \tfrac{\pi}{2})$$



$(x,y) = T^{-1}(r,\varphi) = (r\cos\varphi, r\sin\varphi)$

$$det \left\{ \frac{\partial T^{-1}}{\partial z_j} \right\} = det \left\{ \begin{matrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{matrix} \right\} = r(\cos^2\varphi + \sin^2\varphi) = r$$

$(R, \Phi) = T(X,Y)$ besitzt nach Satz 5.1 eine Dichte
$f_{(R,\Phi)}(r,\varphi) = \frac{1}{2\varphi\sigma^2} e^{-\frac{r^2\cos^2\varphi + r^2\sin^2\varphi}{2\sigma^2}} r \mathbb{1}_{(0,\infty)}(r) \mathbb{1}_{(0,2\pi)}(\varphi) = \frac{r}{\sigma^2} e^{-\frac{-r^2}{2\sigma^2}} \mathbb{1}_{(0,\infty)}(r) \frac{1}{2\pi} \mathbb{1}_{(0,2\pi)}(\varphi)$
Also: R, Φ sind stoch. unabhängig mit
$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \mathbb{1}_{(0,\infty)}(r)$, Rayleigh-Verteilung, $R \sim Ray(\sigma^2)$
$f_\Phi(\varphi) \frac{1}{2\pi} \mathbb{1}_{(0,2\pi)}(\varphi)$, $\Phi \sim N(0, 2\pi)$

Lemma 5.4 $X = (X_1, X_2)$ sei Z'Vektor, absolut stetig mit Dichte $f_X(x_1,x_2)$. Dann ist $Y = X_1 + X_2$ absolut stetig mit Dichte $f_Y(y) = \int\limits_{-\infty}^{\infty} f_X(t, y-t)dt$, $y \in \mathbb{R}$.

Beweis: Mit Satz 5.1[7]: $T(x_1,x_2) = (x_1, x_1+x_2), T^{-1}(y_1,y_2) = (y, y_2 - y_1)$; $det \left\{ \frac{\partial T_i}{\partial x_j} \right\} = det \left\{ \begin{matrix} 1 & 0 \\ 1 & 1 \end{matrix} \right\} = 1$. $f_{T(X)}(y_1,y_2) = f_X(y_1, y_2 - y_1)$. Es folgt: $f_Y(y) = \int\limits_{-\infty}^{\infty} f_X(y_1, y - y_1)dy$. (2. Randverteilung, Integration über 1. Komponente)

Speziell in Lemma 5.4: X_1, X_2 stoch. unabh. $\Longrightarrow f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$
Dann

$$\boxed{f_{X_1+X_2}(y) = \int\limits_{-\infty}^{\infty} f_{X_1}(t) f_{X_2}(y-t)dt \quad , y \in \mathbb{R}}$$

Bei stoch. unabh. ZV'en X_1, X_2 heißt die Verteilung von $X_1 + X_2$ **Faltung der Verteilung** von X_1 und X_2. Bez.: $P^{X_1+X_2} = P^{X_1} * P^{X_2}$ (X_1, X_2 stoch. unabh.)

Beispiel 5.5

a) X_1, X_2 stoch. unabh., $X_1 \sim \Gamma(\alpha, \lambda)$, $X_2 \sim \Gamma(\beta, \lambda)$, Parameter $\alpha, \beta, \lambda > 0$
Dann gilt: $X_1 + X_2 \sim \Gamma(\alpha+\beta, \lambda)$, d.h. $\Gamma(\alpha,\lambda) * \Gamma(\beta,\lambda) = \Gamma(\alpha+\beta,\lambda)$

[7] Wichtiger Trick: Aufblähen der Transformation, damit Satz 5.1 angewendet werden kann.

Beweis: Übung 8

b) X_1, X_2 st. unabh., $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$. Dann gilt: $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, d.h. $N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Beweis: $\int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(t-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(t-\mu_2)^2}{2\sigma_2^2}} dt = ... = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} e^{-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}}$ □

c) $X \sim N(0,1)$. Dann gilt: $X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$

Beweis: Sei $x \geq 0$, $P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq +\sqrt{x}) = \int\limits_{-\sqrt{x}}^{+\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

$\overset{Symmetrie}{=} 2 \int\limits_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u}{2}} \frac{1}{2\sqrt{u}} du$ (Substitution: $u := t^2 \Longrightarrow dt = \frac{1}{2\sqrt{u}} du$)

Damit ist die Dichte von X^2 also: $f_{X^2}(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}} & : x \geq 0 \\ 0 & : sonst \end{cases}$

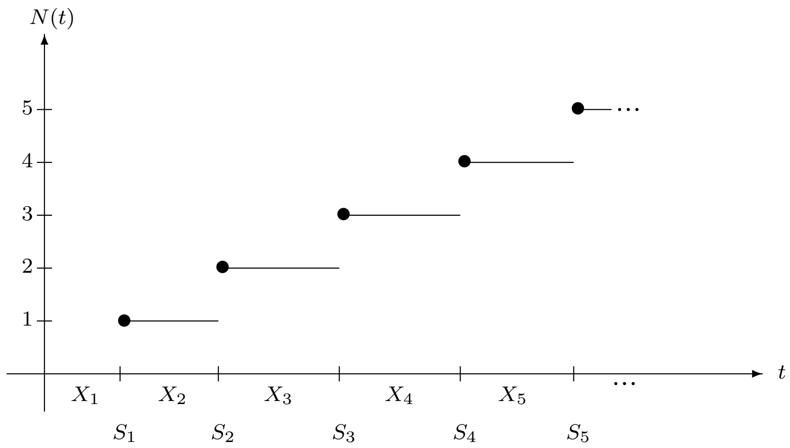
$\frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}$ ist gerade gleich $\frac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} x^{-\frac{1}{2}} e^{-\frac{x}{2}}$ und dies ist die Γ-Verteilung mit Paramtern $\frac{1}{2}$ und $\frac{1}{2}$ (i.z.: $\Gamma(\frac{1}{2}, \frac{1}{2})$) □

Mit a): $X_1, ..., X_n$ *stid* $\sim N(0,1)$. Dann gilt $X_1^2 + ... + X_n^2 \sim \Gamma(\frac{n}{2}, \frac{1}{2})$ Bez.: **χ^2-Verteilung mit n Freiheitsgraden** , χ_n^2
Dichte $f(x) = \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{(0,\infty)}(x)$.
Mit Bsp. 5.3: Sei $Y \sim \chi_2^2 = \Gamma(1, \frac{1}{2}) \Longrightarrow X = \sqrt{Y} \sim Ray(1)$

d) (siehe Bsp. 4.14)
$X_1, ...X_n$ *stid* $\sim Exp(\lambda) = \Gamma(1, \lambda)$, $\lambda > 0$ Mit a): $S_n = X_1 + ... + X_n \sim \Gamma(n, \lambda)$ (**Erlang-Verteilung** mit Parametern n, λ)
Interpretation:
X_i =Bedienzeit eines Servers, *stid* $\sim Exp(\lambda)$
S_n =Gesamtbedienzeit für n Anforderungen.
Umgekehrte Fragestellung: $t \geq 0$. Mit welcher Wahrscheinlichkeit werden im Intervall $[0, t]$ genau n Anforderungen bedient?

Definition 5.6 $\{X_n\}_{n\in\mathbb{N}}$ sei eine Folge von stoch. unabh. ZV'en, $X_i \sim Exp(\lambda)$, $i \in \mathbb{N}$, $\lambda > 0$
$N(t) = max\{n \in \mathbb{N}_0 | \sum\limits_{i=1}^n X_i \leq t\} = |\{n \in \mathbb{N} | \sum\limits_{i=1}^n X_i \leq t\}|$, heißt **Poisson-Prozeß** mit Paramter $\lambda > 0$ (Bez. $PP(\lambda)$). $N(t)$ ist die Anzahl der bis zur Zeit $t \geq 0$ bedienten Kunden.

Grafisch:



Sprechweisen:
X_n heißen **Zwischenankunftszeit**, **Verweilzeiten** (interarrival times, sojourn times)
$S_n = \sum\limits_{i=1}^{n} X_i$ heißen **Ankunftszeiten** (arrival times)

Lemma 5.7 $\forall t \geq 0$ besitzt $N(t)$ eine Poissonverteilung mit Parameter λt, d.h. $N(t) \sim Poi(\lambda t)$, $P(N(t) = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$, $k \in \mathbb{N}_0$.

Beweis:

(i) $P(N(t) = 0) = P(X_1 > t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t} = e^{-\lambda t}\frac{(\lambda t)^0}{0!}$

(ii) $P(N(t) = k) = P(\underbrace{\sum\limits_{i=1}^{k} X_i}_{S_k} \leq t, \underbrace{\sum\limits_{i=1}^{k+1} X_i}_{S_{k+1}} > t) = P(\{S_k \leq t\} \backslash \{S_{k+1} \leq t\})$[8]

$\overset{Monotonie\ von\ P}{=} P(S_k \leq t) - P(S_{k+1} \leq t)$

Es ist: $S_k \sim Erl(k, \lambda)$, $S_{k+1} \sim Erl(k+1, \lambda)$ Somit folgt:

$P(S_k \leq t) - P(S_{k+1} \leq t) = \frac{\lambda^k}{(k-1)!}\int\limits_0^t x^{k-1}e^{-\lambda x}dx - \frac{\lambda^{k+1}}{k!}\int\limits_0^t x^k e^{-\lambda x}dx$

$= \frac{\lambda^k}{(k-1)!}(\int\limits_0^t x^{k-1}e^{-\lambda x}dx - \frac{\lambda}{k}\int\limits_0^t x^k e^{-\lambda x}dx)$

$= \frac{\lambda^k}{(k-1)!}(\int\limits_0^t \underbrace{x^{k-1}}_{u'}\underbrace{e^{-\lambda x}}_{v}dx + \int\limits_0^t \underbrace{\frac{x^k}{k}}_{u}\underbrace{(-\lambda e^{-\lambda x})}_{v'}dx)$

Mit $\int uv' = uv - \int u'v$ folgt dann:[9]

$\frac{\lambda^k}{(k-1)!}\frac{x^k}{k}e^{-\lambda x}|_0^t = \frac{\lambda^k}{k!}t^k e^{-\lambda t} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$ □

Es gilt sogar für $N_{(s,t]} = N(t) - N(s)$ (Anzahl der Kunden im Intervall $(s, t]$), $0 \leq s < t$, **Zuwachs** in $(s, t]$:

[8] $A \cap B = A \backslash \overline{B}$
[9] Partielle Integration

- $N_{(s,t]} \sim Poi(\lambda(t-s))$
- $N_{(s_i,t_i]}$, $i \in \mathbb{N}$, sind stoch. unabh., $Poi(\lambda(t_i - s_i))$-verteilt, falls die Intervalle $(s_i, t_i]$ p.d., d.h. die Zuwächse eines PP sind stoch. unabh. poissonverteilte ZV'en.

Im Folgenden: Summen von diskreten ZV'en.

Lemma 5.8 X_1, X_2 stoch. unabh., diskret mit Träger $\mathbb{N}_0$, Zähldichten f_{X_1}, f_{X_2}. Dann besitzt $X_1 + X_2$ die Zähldichte

$$f_{X_1+X_2}(k) = \sum_{i=0}^{k} f_{X_1}(i) f_{X_2}(k-i), \quad k \in \mathbb{N}_0$$

Beweis: $P(X_1 + X_2 = k) = P(\bigcup_{i=0}^{k}(\{X_1 = i\} \cap \{X_2 = k-i\}))$

$= \sum_{i=0}^{k} P(X_1 = i, X_2 = k-i) \overset{st.u.}{=} \sum_{i=0}^{k} \underbrace{P(X_1 = i)}_{f_{X_1}(i)} \underbrace{P(X_2 = k-i)}_{f_{X_2}(k-i)}$ □

Beispiel 5.9 $X_1, ..., X_n$ *stid* $\sim$ *Geo*(p), $0 \leq p < 1$. Dann gilt

$$P(\sum_{i=1}^{n} X_i = k) = \binom{n+k-1}{n-1}(1-p)^k p^n, \ k \in \mathbb{N}_0.$$ [10]

Diese Verteilung heißt **negative Binomialverteilung**[11], Bez. $\overline{Bin}(n,p)$. Also: $\underbrace{Geo(p) * ... * Geo(p)}_{\text{n-mal}} = \overline{Bin}(n,p)$

$X_1 + ... + X_n$ = Wartezeit bis zum n-ten Treffer (ohne die Treffer mitzuzählen)

Lemma 5.10

a) $Bin(n_1, p) * Bin(n_2, p) = Bin(n_1 + n_2, p), \ n_1, n_2 \in \mathbb{N}, \ 0 \leq p < 1$. Insbesondere gilt: $\underbrace{Bin(1,p) * ... * Bin(1,p)}_{n-mal} = Bin(n,p)$

b) $\overline{Bin}(n_1, p) * \overline{Bin}(n_2, p) = \overline{Bin}(n_1 + n_2, p), \ n_1, n_2 \in \mathbb{N}, \ 0 < p < 1$

c) $Poi(\lambda_1) * Poi(\lambda_2) = Poi(\lambda_1 + \lambda_2)$

Diese 3 Verteilungsklassen sind **faltungsstabil**

[10] Mit $\binom{n}{k} = \binom{n}{n-k}$ folgt $\binom{n+k-1}{n-1}(1-p)^k p^n = \binom{n+k-1}{k}(1-p)^k p^n$

[11] Die negative Binimialverteilung wird auch Pascalverteilung genannt.

Lemma 5.11 X_1, X_2 stoch. unabh., abs.-stetig mit Dichten $f_{X_1}, f_{X_2},\ f_{X_i}(x) = 0$, falls $x \leq 0,\ i = 1, 2$. dann gilt:

a) $Y = X_1 X_2$ ist abs.-stetig mit Dichte $f_Y(y) = \int\limits_0^\infty \frac{1}{t} f_{X_1}(\frac{y}{t}) f_{X_2}(t) dt\ \mathbb{1}_{(0,\infty)}(x)$[12]

b) $Z = \frac{x_1}{x_2}$ ist abs.-stetig mit Dichte $f_Z(y) = \int\limits_0^\infty t f_{X_1}(yt) f_{X_2}(t) dt\ \mathbb{1}(0, \infty)(x)$

Beweis: Transformationssatz für Dichten mit $T(x, y) = (x, xy)$ bzw. $T(x, y) = (x, \frac{x}{y})$ $\square$

[12]Benutze Transformation: $T(x, y) = (y, xy)$ (Aufblähen; siehe Beweis Lemma 5.4)

6 Erwartungswerte und Momente von ZV'en

Zur Motivation 2 Beispiele

(i) Spiel: symmetrischer Würfel
Modell: ZV'e X auf einem W'Raum $(\Omega, \mathfrak{A}, P)$, $P(X=i) = \frac{1}{6}$, $i = 1, ..., 6$

- Auszahlung = Augenzahl in Euro
  Frage: 'mittlere' Auszahlung, erwartete Auszahlung (=Einsatz beim Spiel)
  $E(X) = \frac{1}{6}1 + \frac{1}{6}2 + ... + \frac{1}{6}6 = \frac{1}{6}(1 + ... + 6) = 3,5$
- Auszahlung = Augenzahl zum Quadrat (in Euro)
  $g(X) = X^2$
  $E(g(X)) = \frac{1}{6}1^2 + \frac{1}{6}2^2 + ... + \frac{1}{6}6^2 = \frac{1}{6}(1^2 + 2^2 + ... + 6^2) = 15\frac{1}{6}$

(ii) Roulette (Petersburger Paradoxon) (Nikolaus Bernoulli, 1695-1726)
Strategie: Beim n-ten Spiel setzte 2^n Euro auf 'rot', $n \in \mathbb{N}_0$. Bei Gewinn $\rightarrow$ STOP, Verlust: $n \rightarrow n+1$
Erstmalig 'rot' beim n-ten Spiel bedeutet:
Gesamteinsatz: $1 + 2 + 4 + 8 + ... + 2^n = 2^{n+1} - 1$
Auszahlung: $= 2^{n+1}$
Gewinn: $= 1$ Euro

Aber ist die eine sichere Gewinnstrategie?
Modell: $\{X_n\}_{n \in \mathbb{N}_0}$ *stid*, $X \sim Bin(1, \frac{1}{2})$, $\{X_n = 1\}$: 'rot' im n-ten Spiel.
$S = min\{n \in \mathbb{N}_0 | X_n = 1\}$: Zeitpunkt von erstmalig 'rot'; $S \sim Geo(\frac{1}{2})$, $P(S=k) = (1-p)^k p$, $n \in \mathbb{N}_0$
Auszahlung: $A = 1$, falls $S < \infty$, $P(A=1) = P(S < \infty) = \sum\limits_{k=0}^{\infty} P(S = k) = 1$. Erwartete Auszahlung: $E(A) = 1P(A=1) = 1$
Ist eine sichere Gewinnstrategie, erfordert allerdings unendliches Kapital und auch unendliche Zeit.
Ein genaueres Modell wäre:
Annahme: max. Kapital= $2^L - 1$ Euro $\Longrightarrow$ höchstens $L-1$ Spiele spielbar
Auszahlung: $A = 1 \Longleftrightarrow S \leq L-1$ bzw. $A = -(2^L - 1) \Longleftrightarrow S \geq L$
$P(A=1) = P(S \leq L-1) = \sum\limits_{k=0}^{L-1} \frac{1}{2^{k+1}} = 1 - \frac{1}{2^L}$
$P(A = -(2^L - 1)) = P(S \geq L) = 1 - (P(A=1)) = \frac{1}{2^L}$
Erwartete Auszahlung: $E(A) = 1(1 - \frac{1}{2^L}) - (2^L - 1)\frac{1}{2^L} = 0$

In beiden Beispielen: $E(A) = \sum\limits_i iP(X=i)$ bzw.

$$E(g(X)) = \sum_i g(i)P(X=i), \quad i \text{ Trägerpunkte}$$

Beachte bei der Erweiterung von E auf ∞ viele Trägerpunkte oder als abs-stetige ZV'e, daß Σ bzw. $\int$ wohldefiniert sind.

Definition 6.1 (Erwartungswerte von ZV'en)
g sei eine reellwertige Funktion.

a) X sei diskrete ZV'e mit Träger $T = \{x_1, x_2, ...\} \subset \mathbb{R}$ und Zähldichte f. Falls $\sum\limits_{i=1}^{\infty} |g(x_i)| f(x_i) < \infty$, so heißt $E(g(X)) = \sum\limits_{i=1}^{\infty} g(x_i) f(x_i) = \sum\limits_{i=1}^{\infty} g(x_i) P(X = x_i)$ **Erwartungswert** von $g(X)$.

b) X sei absolut-stetig mit Dichte f. Falls $\int\limits_{-\infty}^{\infty} |g(x)| f(x) dx < \infty$, so heißt $E(g(X)) = \int\limits_{-\infty}^{\infty} g(x) f(x) dx$ **Erwartungswert** von $g(X)$.

Insbesondere $g(X) = x$ (Identität):
$E(X) = \sum\limits_{i=1}^{\infty} x_i P(X = x_i)$ bei diskrete ZV'en und
$E(X) = \int\limits_{-\infty}^{\infty} x f(x) dx$ bei abs.-stetigen ZV'en
Für beliebige ZV'e kann der Erwartungswert mit Hilfe der Verteilungsfunktion wie folgt berechnet werden.

Lemma 6.2 X sei ZV'e mit Verteilungsfunktion F. Falls $\int\limits_{-\infty}^{0} F(x) dx < \infty$ und $\int\limits_{0}^{\infty} (1 - F(x)) dx < \infty$, so gilt:

$$E(X) = -\int\limits_{-\infty}^{0} F(x) dx + \int\limits_{0}^{\infty} (1 - F(x)) dx$$

Beweis: Nur für den Fall differenzierbarer Verteilungsfunktionen ($F'(x) = f(x)$ ist eine Dichte)

$$\int\limits_{-\infty}^{0} F(x) dx = \int\limits_{-\infty}^{0} \underbrace{F(x)}_{u} \underbrace{1}_{v'} dx = F(x) x|_{-\infty}^{0} - \int\limits_{-\infty}^{0} x f(x) dx = -\int\limits_{-\infty}^{0} x f(x) dx$$

$$\int\limits_{0}^{\infty} (1-F(x)) dx = \int\limits_{0}^{\infty} \underbrace{(1 - F(x))}_{u} \underbrace{1}_{v'} dx = (1-F(x)) x|_{0}^{\infty} - \int\limits_{0}^{\infty} -f(x) x dx = \int\limits_{0}^{\infty} f(x) x dx$$

Also $E(X) = \int\limits_{-\infty}^{0} x f(x) dx \int\limits_{0}^{\infty} x f(x) dx = -\int\limits_{-\infty}^{0} F(x) dx + \int\limits_{0}^{\infty} (1 - F(x)) dx$ □

Beispiel 6.3

a) Geometrische Verteilung, $X \sim Geo(p)$, $P(X = k) = f(k)$
$= (1 - p)^k p$, $k \in \mathbb{N}_0$, $0 < p \leq 0$ (keine Degenerierung)

$E(X) = \sum\limits_{k=0}^{\infty} k(1-p)^k p = \sum\limits_{k=1}^{\infty} k(1-p)^k p \stackrel{!}{=} \sum\limits_{k=0}^{\infty} (k+1)(1-p)^{k+1} p$
$= (1-p) \underbrace{\sum_{k=0}^{\infty} k(1-p)^k p}_{E(X)} + (1-p) \underbrace{\sum_{k=0}^{\infty} (1-p)^k p}_{1}$
Also: $E(X) = (1-p)E(X) + (1-p) = E(X) - pE(X) + 1 - p \Longrightarrow p(EX) = 1 - p$
und somit: $p \sum\limits_{k=0}^{\infty} k(1-p)^k p = 1 - p$
$E(X) = \frac{1-p}{p}$, $o < p \leq 1$

b) Exponentialverteilung, $X \sim Exp(\lambda)$, $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$ $(\lambda > 0)$
$E(X) = \int\limits_0^{\infty} \underbrace{x}_{u} \underbrace{\lambda e^{-\lambda x}}_{v'} dx = -xe^{-\lambda x}|_0^{\infty} + \int\limits_0^{\infty} e^{-\lambda x} dx = -\frac{1}{\lambda} e^{-\lambda x}|_0^{\infty} = \frac{1}{\lambda}$
Also: $E(X) = \frac{1}{\lambda}$, falls $X \sim Exp(\lambda)$

c) Normalverteilung, $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$, $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in \mathbb{R}$
$E(X) = \int\limits_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} (x+\mu) e^{-\frac{x^2}{2\sigma^2}} dx$
$= \frac{1}{\sqrt{2\pi}\sigma} \underbrace{\int_{-\infty}^{\infty} x e^{-\frac{x^2}{2\sigma^2}} dx}_{0} + \mu \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx}_{1} = \mu$

Lemma 6.4 *X sei diskrete ZV'e mit Träger $\mathbb{N}_0$. Dann gilt*

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k)$$

Beweis: Mit Lemma 6.2:
$E(X) = \int\limits_0^{\infty} (1 - F(x)) dx = \sum\limits_{k=0}^{\infty} P(X > k) = \sum\limits_{k=1}^{\infty} P(X \geq k)$ □

Satz 6.5 (Eigenschaften des Erwartungswertes) [13]
(auftretende EW sollen existieren)

a) $E(aX + bY) = aE(X) + bE(Y)$ $\quad \forall a, b \in \mathbb{R}$ (Linearität)
b) $X \leq Y \Longrightarrow E(X) \leq E(Y)$ (Monotonie)
c) Für $X = \mathbb{1}_A$, $A \in \mathfrak{A}$, gilt $E(X) = E(\mathbb{1}_A) = P(A)$
d) $P(|X| > c) \leq \frac{E(|X|)}{c}$ $\quad \forall c > 0$ (Markoff-Ungleichung)
e) X, Y stoch. unabh. $\Longrightarrow E(XY) = E(X)E(Y)$

[13] Im Folgenden wird manchmal für E(X) auch E[X] oder EX geschrieben, diese 3 Bez. sind äquivalent

Für den Beweis wird benötigt $E(X)$, wobei $Z = aX + b$
Erwartungwert von ZV'en werden wie folgt berechnet:

Satz 6.6 $(X_1, ..., X_n)$ sei Z'Vektor. $g : \mathbb{R}^n \to \mathbb{R}^1$ meßbare Funktion

a) $(X_1, ..., X_n)$ sei diskret mit Träger $T = \{t_1, t_2, ...\} \subset \mathbb{R}^n$ und Zähldichte f.
Falls $\sum_{i=0}^{\infty} |g(t_i)| f(t_i) < \infty$, so gilt: $E(g(X_1, ..., X_n)) = \sum_{i=0}^{\infty} g(t_i) f(t_i)$.

$\boxed{E(X,Y) = \sum_i \sum_j x_i y_j f_{(X,Y)}(x_i, y_i) \qquad \text{siehe unten im Beweis von 6.5 e)}}$

b) $(X_1, ..., X_n)$ sei abs.-stetig mit Dichte f.
Falls $\int \cdots \int |g(x_1, ..., x_n)| f(x_1, ..., x_n) dx_1 \cdots dx_n < \infty$, so gilt:
$E(g(X_1, ..., X_n)) = \int \cdots \int g(x_1, ..., x_n) f(x_1, ..., x_n) dx_1 \cdots dx_n$

Beweis: (von Satz 6.5)

a) Für abs.-stetige ZV:
$E(aX + bY) = \int \int (ax + by) f_{(x,y)}(x, y) dx dy$
$= a \int \int x f_{(X,Y)}(x, y) dx dy + b \int \int y f_{(X,Y)}(x, y) f x f y$
$= a \underbrace{\int x f_X(x) dx}_{E(X)} + b \underbrace{\int y f_Y(y) dy}_{E(Y)} = aE(X) + bE(Y)$

b) Übung

c) $\mathbb{1}_A(\omega) = \begin{cases} 1 & : \omega \in A \\ 0 & : \omega \notin A \end{cases}$ somit $E(\mathbb{1}_A) = 1P(A) + 0P(\overline{A}) = P(A)$

d) $\forall\, c > 0 :\ c \cdot \mathbb{1}_{|X|>c} = \left\{ \begin{array}{ll} c & : |X| > c \\ 0 & : |X| \leq c \end{array} \right\} \leq |X|$
$E(c \cdot \mathbb{1}_{|X|>c}) = c \cdot P(|X| > c) \leq E(X)$ Also folgt die Behauptung.

e) nur für diskrete ZV (stetiger Fall analog)
$E(XY) = \sum_{i,j} x_i y_i f_{(X,Y)}(x_i, y_j) \overset{s.u.}{=} \sum_{i,j} x_i y_i f_X(x_i) f_Y(y_j)$
$= \underbrace{(\sum_i x_i f_X(x_i))}_{E(X)} \underbrace{(\sum_j y_j f_Y(y_j))}_{E(Y)} = E(X)E(Y)$ $\square$

Definition 6.7 X, Y seien ZV, X und Y heißen **unkorreliert**, wenn $E(XY) = E(X)E(Y)$

Also mit Satz 6.5 e): X, Y s.u. $\Longrightarrow X, Y$ unkorreliert. Die Umkehrung hiervon ist im Allg. falsch.
Transformationen von besonderer Bedeutung sind:

$$g(X) = X^k,\ k \in \mathbb{N}$$

und

$$g(X,Y) = (X - EX)(Y - EY)$$

Definition 6.8 X, Y seien ZV, alle auftretenden E-Werte sollen existieren.

a) $E(X^k),\ k \in \mathbb{N}_0$, heißt **k-tes Moment** von X (k=1: E-Wert von X)

b) $E((X - EX)^k)$ heißt **k-tes zentrales Moment**
Insbes. $k = 2$: $E((X - EX)^2) = V(X) = Var(X)$ heißt **Varianz**[14] von X
$\sqrt{Var(X)}$ heißt **Standartabweichung** von X

c) $Cov(X,Y) = E((X - EX)(Y - EY))$ heißt **Kovarianz** von X und Y.
$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$ heißt **Korrelation** von X und Y.

Lemma 6.9 $X, Y, X_1, ..., X_n$ seien ZV, auftretende Moment sollen existieren.

a) $Cov(X,Y) = E(XY) - E(X)E(Y)$
$Cov(X,X) = Var(X)$, $Var(X) = E(X^2) - (E(X))^2$

b) $Var(aX + b) = a^2 Var(X)\ \forall a, b \in \mathbb{R}$

c) $Var(\sum\limits_{i=1}^{n} X_i) = \sum\limits_{i=1}^{n} Var(X_i) + 2\sum\limits_{i<j} Cov(X_i, X_j)$
$Var(\sum\limits_{i=1}^{n} X_i) = \sum\limits_{i=1}^{n} Var(X_i)$, falls $X_1, ..., X_n$ p.w. unkorreliert.

d) $|Cov(X,Y)| \leq \sqrt{Var(X)Var(Y)}$ (Cauchy-Schwarz-Ungleichung)
Insbesondere folgt: $|Corr(X,Y) \leq 1|$

Beweis:

a) $Cov(X,X) = E[(X - EX)(X - EX)] = E[(X - EX)^2] = Var(X)$

$Cov(X,Y) = E[(X - EX)(Y - EY)] = E[XY - X(EY) - Y(EX) + (EX)(EX)] = E(XY) - (EY)(EX) - (EX)(EY) + (EX)(EY) = E(XY) - (EX)(EY)$

$Var(X) = Cov(X,X) = E(X^2) - (EX)^2$

b) $Var(aX + b) = E[((aX + b) - E(aX + b))^2] = E(aX + b - a(EX) - b)^2 = a^2 E[(X - EX)^2] = a - Var(X)$

c) $Var(\sum\limits_{i=1}^{n} X_i) = E[(\sum\limits_{i=1}^{n} X_i)^2] - (E(\sum\limits_{i=1}^{n}))^2 = \sum\limits_{i,j} E(X_i X_j) - \sum\limits_{i,j}(EX_i)(EX_j) = \sum\limits_{i} Var(X_i) + 2\sum\limits_{i<j} E(X_i X_j) - (EX_i)(EX_j) = \sum\limits_{i} Var(X_i) + 2\sum\limits_{i<j} Cov(X_i, X_j)$

[14]Die Varianz wird auch als Streuungsmaß bezeichnet

d) Hierbei die Existenz von Dichten
$|E(XY)| = |\int\int xyf_{(X,Y)}dxdy| \leq \int\int |x||y|f_{(X,Y)}dxdy \overset{Cauchy-Schw.-Ungl.}{\leq}$
$\sqrt{(\int\int x^2 f(x,y)dxdy)(\int\int y^2 f(x,y)dxdy)} = \sqrt{(\int x^2 f_X(x)dx)(\int y^2 f_X(x)dx)}$
$= \sqrt{(EX^2)(EY^2)}$
Die Behauptung erfolgt durch Ersetzen: $X \leftarrow X - EX$ und $Y \leftarrow Y - EY$

□

Interpretation der Größen aus Lemma 6.9:

$E(X)$	: mittlerer Wert
$Var(X)$	: mittlere quadratische Abweichung vom Erwartungswert, Maßzahl für die Streuung einer ZV um den E-Wert
$Cov(X,Y)$	: Korrekturterm bei der Varianzberechnung von Summen von ZV. $Var(X,Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
$Corr(X,Y)$	: Maßzahl für den linearen Zusammenhang zwischen ZV. Es gilt: $\exists a, b \in \mathbb{R} : P(X = aY + b) = 1 \iff |Corr(X,Y)| = 1$

Satz 6.10 $G_{X_1}(z), G_{X_2}(z)$ seien erzeugende Funktionen von s.u. diskreten ZV X_1, X_2, bzw. $L_{X_1}(s), L_{X_2}(s)$ die LaPlace-Transformierten von s.u. abs.-stetigen ZV $X_1, X_2 \geq 0$. Dann gilt:

$$G_{X_1+X_2}(z) = G_{X_1}(z)G_{X_2}(z),\ |z| \leq 1$$

bzw

$$L_{X_1+X_2}(s) = L_{X_1}(s)L_{X_2}(s),\ s \geq 0$$

Beweis: Es gilt:
$G_X(z) = \sum\limits_{k=0}^{\infty} p_k z^k = E(z^k)$ und $L_X(s) = \int\limits_0^{\infty} e^{-sx}f(x)dx = E(e^{-sx})$ Also:
$G_{X_1+X_2} = E(z^{X_1+X_2}) = E(z^{X_1}z^{X_2}) \overset{s.u.}{=} E(z^{X_1})E(z^{X_2}) = G_{X_1}(z)G_{X_2}(z),\ |z| \leq 1$

Und im stetigen Fall gilt dann:
$L_{X_1+X_2}(s) = E(e^{-s(X_1+X_2)}) = E(e^{-sX_1}e^{-sX_2}) \overset{s.u.}{=} E(e^{-sX_1})E(e^{-sX_2})$
$= L_{X_1}(s)L_{X_2}(s),\ s \geq 0$

Beispiel 6.11

a) $X \sim Bin(n,p)$. Dann $G_X(z) = (1-p+pz)^n$, $|z| \leq 1$. $X_1 \sim Bin(n_1,p)$, $X_2 \sim Bin(n_2,p)$ s.u. $\Longrightarrow G_{X_1+X_2}(z) = (1-p+pz)^{n_1}(1-p+pz)^{n_2} = (1-p+pz)^{n_1+n_2}$
Dies ist gerade die erzeugende Funktion einer $Bin(n_1+n_2,p)$-Verteilung.

b) $X_1 \sim Exp(\lambda),\ L_{X_1}(s) = \frac{\lambda}{s+\lambda},\ s \geq 0$
$X_2 \sim Exp(\mu),\ L_{X_2}(s) = \frac{\mu}{s+\mu},\ s \geq 0$
$L_{X_1+X_2}(s) = \frac{\lambda}{s+\lambda}\frac{\mu}{s+\mu} = \frac{\mu}{\mu-\lambda}\frac{\lambda}{s+\lambda}\frac{\lambda}{\lambda-\mu}\frac{\lambda}{s+\lambda}$ (1)
Die Dichte $f(x) = (\frac{\mu}{\mu-\lambda}\lambda e^{-\lambda x} + \frac{\lambda}{\lambda-\mu}\mu e^{-\mu x})\mathbb{1}_{[0,\infty)}(x)$ (2)
hat LaPlace-Transformierte (1); $X_1 + X_2$ besitzt die Dichte (2)

Satz 6.12 (Transformierte und Momente)

a) $G(z) = \sum\limits_{k=0}^{\infty} p_k z^k = \sum\limits_{k=0}^{\infty} P(X = k) z^k$ sei erzeugende Funktion einer diskreten ZV X mit Träger $\mathbb{N}_0$. Dann gilt:

$$E(X) = G'(1),\ E(X(X-1)...(X-k+1)) = G^{(k)}(1)$$

$G^{(k)}(1)$ ist das **k-te faktorielle Moment**.
Also: $E(x^2) = G'(1) + G''(1)$; Beachte: $G^{(i)}(1) = \lim\limits_{z \uparrow 1} G^{(i)}(z)$, falls $G(z)$ nicht ex. für $z > 1$
$\boxed{Var(X) = G'(1) + G''(1) - (G'(1))^2}$

b) $L(s) = \int\limits_0^{\infty} e^{-sx} f(x) dx$ sei LaPlace-Transformierte einer abs.-stetigen ZV $X > 0$ mit Dichte f. Dann gilt:

$$E(X) = -L'(0),\ E(X^k) = (-1)^k L^{(k)}(0)$$

$\boxed{Var(X) = L''(0) - (L'(0))^2}$

Beweis: Nur a), E(X):
$G'(z) = \sum\limits_{k=0}^{\infty} p_k k z^{k-1},\ |z| < 1$ (gliederweise Differenzieren).
Also: $G'(1) = \sum\limits_{k=0}^{\infty} p_k k = E(X)$ □

Beispiel 6.13 (Berechnungen von Momenten)

a) $X \sim Bin(n, p),\ 0 \leq p \leq 1$, Berechnung des Erwartungswertes:

1. $EX = \sum\limits_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = ... = np$
2. $X \sim \sum\limits_{i=1}^{n} X_i,\ X_i stid \sim Bin(1, p)$
   $E(X_i) = 1p + 0(1-p) = p$
   $E(X) = E(\sum\limits_{i=1}^{n} X_i) = \sum\limits_{i=1}^{n} (E(X_i)) = \sum\limits_{i=1}^{n} p = np$
3. Erzeugende Funktion
   $G(z) = (pz + 1 - p)^n$
   $G'(z) = n(1 + pz - z)^{n-1} p$
   $G'(1) = np = E(X)$

$Var(X)$ wie in 2):
$Var(X) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$
Es ist: $Var(X_i) = E(X_i^2) - (E(X_i))^2 = E(X_i^2) - p^2$
$E(X_i^2) = 1^2 p + 0^2(1-p) = p$
Also $Var(X_i) = p - p^2 = p(1-p)$
Somit: $\sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$

b) $X \sim Exp(\lambda),\ \lambda > 0$
$L_X(s) = \frac{\lambda}{s+\lambda}$
$L_X^{(k)}(s) = (-1)^k k! \frac{\lambda}{(s+\lambda)^{k+1}}$
$L_X^{(k)}(0) = (-1)^k k! \lambda^{-k}$
Also: $E(X^2) = k!\lambda^{-k},\ E(X) = \frac{1}{\lambda},\ E(X^2) = \frac{2}{\lambda^2}$
$Var(X) = E(X^2) - (E(X))^2 = \frac{2}{\lambda^2}\frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

c) $X \sim N(0,1)$, d.h. normalverteilte ZV mit $\mu = 0,\ \sigma^2 = 1$
Dann $E(X) = 0$
$E(X^2) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int\limits_{0}^{\infty} x^2 e^{-\frac{x^2}{2}} dx$
(Substitution: $\frac{x^2}{2} =: y,\ dx = \frac{dy}{\sqrt{2y}}$)

$$= \frac{2}{\sqrt{2\pi}} \int\limits_0^\infty 2ye^{-y} \frac{1}{\sqrt{2y}} dy = \underbrace{\frac{2}{\sqrt{\pi}} \int\limits_0^\infty y^{\frac{1}{2}} e^{-y} dy}_{\text{Integral über die Dichte einer } \Gamma(\frac{3}{2},1)-\text{Verteilung}} = 1$$

Es folgt: $Var(X) = E(X^2) - (EX)^2 = 1 - 0 = 1$ (Varianz einer Standart-Normalverteilung ist 1)
$Y = \sigma X + \mu$ besitzt die Dichte $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, d.h. $Y \sim N(\mu, \sigma^2)$
Also: $E(Y) = E(\sigma X + \mu) = \sigma \underbrace{E(X)}_{=0} + \mu = \mu$
$Var(Y) = Var(\sigma X + \mu) = \sigma^2 \underbrace{Var(X)}_{=1} ? \sigma^2$
Also: Bei $N(\mu, \sigma^2)$ ist μ der Erwartungswert und σ^2 die Varianz.

7 Bedingte Verteilungen und Erwartungswerte

a) <u>diskreter Fall</u>
X, Y diskrete ZV mit Träger T_X, T_Y, gemeinsamer Dichte $f_{X,Y} = P(X = x, Y = y),\ (x, y) \in T_X \times T_Y$
Def. **bedingte Zähldichte von** X **unter** $Y = y$ (analog zur Def. 2.15)

$$f_{X|Y}(x|y) = \begin{cases} \frac{P(X=x,Y=y)}{P(Y=y)} = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} & : f_Y(y) > 0 \\ f_X(x) \text{ (oder beliebige andere Zähldichte)} & : f_Y(y) = 0 \end{cases}$$

Beachte: $f_{X|Y}(x|y)$ <u>ist</u> eine Zähldichte $\forall y \in T_y$ fest. Die zugehörige Verteilung heißt **bedingte Verteilung von** X **unter** $Y = y$, also $P(X \in A|Y = y) = P^{X|Y=y}(A) = \sum\limits_{x \in A} f_{X|Y}(x|y),\ A \in \mathfrak{A},\ y \in T_Y$.
Satz von der totalen W'keit (s. Satz 2.15) liefert :
$f_X(x) = P(X = x) = \sum\limits_{y \in T_Y} f_{X|Y}(x|y) f_Y(y),\ x \in T_X$ $\qquad(\star)$

Beispiel 7.1 Seien $f_{X|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, ..., n,$ also $P^{X|N=n} = Bin(n, p)$ und $N \sim Poi(\lambda)$.
2 stufiges Experiment:

$$\begin{aligned} P(X = k) &= \sum_{n=0}^{\infty} P(X = k|N = n)P(N = n) = \sum_{n=k}^{\infty} P(X = k|N = n)P(N = n) \\ &= \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \frac{\lambda^k p^k}{k!} \underbrace{\sum_{n=k}^{\infty} \frac{(\lambda(1-p)^{n-k})}{(n-k)!}}_{e^{\lambda(1-p)}} \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda p} \sim Poi(\lambda p) \end{aligned}$$

d.h.: $X \sim Poi(\lambda p)$

Betrachte Erwartungswert bzgl. der bedingten Verteilung:
$E(g(X)|Y = y) = \sum\limits_{x \in T_X} g(x) f_{X|Y}(x|y),\ y \in T_Y$, heißt **bedingter Erwartungswert von** $g(x)$ **unter** $Y = y$ (falls existent)
Berechnung des Erwartungswertes mit den bedingten Erwartungswerten wie folgt:

$$\begin{aligned} E(g(X)) &= \sum_{x \in T_X} g(x) f_X(x) = \sum_{x \in T_X} g(x) \sum_{y \in T_Y} f_{X|Y}(x|y) f_Y(y) \\ &= \sum_{y \in T_Y} \underbrace{\Big(\sum_{x \in T_X} g(x) f_{X|Y}(x|y) \Big)}_{E(g(X)|Y=y)} f_Y(y) = \sum_{y \in T_Y} E(g(X)|Y = y) f_Y(y) \end{aligned}$$

b) absolut-stetiger Fall
X, Y abs.-stetige ZV mit gemischter Dichte $f_{(X,Y)}(x,y)$.
Def. **bedingte Dichte von X unter $Y = y$** (analog zur Def. 2.15) durch:

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & : f_Y(y) > 0 \\ f_X(x) \text{ (oder beliebige andere Dichte)} & : f_Y(y) = 0 \end{cases}$$

Beachte: $f_{X|Y}(x|y)$ ist eine Dichte $\forall y \in \mathbb{R}$. Die zugehörige Verteilung heißt **bedingte Verteilung von X unter $Y = y$**, also $P(X \in A|Y = y) = P^{X|Y=y}(A) = \int\limits_A f_{X|Y}(x|y)dx, \ A \in \mathfrak{A}, \ y \in \mathbb{R}$.
Speziell heißt:
$F_{X|Y=y}(x) = P(X \le x|Y = y) = \int\limits_{-\infty}^{\infty} f_{X|Y}(z|y)dz$ **bedingte Verteilungsfunktion von X unter $Y = y$**. Wie bei A) $(\star)$: $f_X(x) = \int\limits_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy$
Analog zum diskreten Fall
$E(g(X)|Y = y) = \int g(x)f_{X|Y}(x|y)dx, \ y \in \mathbb{R}$ heißt **bedingter Erwartungswert von $g(x)$ unter $Y = y$**

$$\begin{aligned} E(g(X)) &= \int g(x)f_X(x)dx = \int g(x) \int f_{X|Y}(x|y)f_Y(y)dydx \\ &= \int \underbrace{\left(\int g(x)f_{X|Y}(x|y)dx \right)}_{E(g(X)|Y=y)} f_Y(y)dy = \int E(g(X)|Y = y)f_Y(y)dy \end{aligned}$$

c) gemischte Fälle (abs.-stetig/diskret werden analog behandelt)

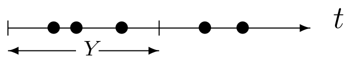
Beispiel 7.2 (**Wartezeit in Warteschlangensystem**) N: Anzahl der Kunden in der Warteschlange, $N \sim Geo(p)$
$X_n, n \in \mathbb{N}$: Bedienzeit von Kunde n, $X_n \ stid \sim Exp(\lambda)$
Gesucht: $W = \sum\limits_{i=1}^{N+1} X_i$: Gesamtwartezeit für neu ankommende Kunden.
Bekannt: $P^{W|N=n} = Erl(n+1, \lambda)$. Es gilt:

$$\begin{aligned} P(W \le t) &= \sum_{n=0}^{\infty} P(W \le t|N = n)P(N = n) \\ &= \sum_{n=0}^{\infty} \int\limits_0^t \frac{\lambda^{n+1}}{n!} y^n e^{-\lambda y} dy (1-p)^n p = \int\limits_0^t \lambda p e^{-\lambda y} \sum_{n=0}^{\infty} \frac{(\lambda y(1-p))^n}{n!} dy \\ &= \int\limits_0^t \lambda p e^{-\lambda y} e^{\lambda y(1-p)} dy = \int\limits_0^t \lambda p e^{-\lambda p y} dy = 1 - e^{-\lambda p t}, \ t \ge 0 \end{aligned}$$

Also: $W \sim Exp(\lambda p)$ mit $E(W) = \frac{1}{\lambda p}$

Beispiel 7.3 (Ankünfte eines PP in zufälligen Intervallen)

$\left.\begin{array}{l} N(t) : PP,\ \lambda > 0 \\ Y \sim Exp(\mu),\ \mu > 0 \end{array}\right\}$ s.u.



Gesucht: Anzahl der Ankünfte in einem zufälligen Intervall $[0, Y]$, $N(Y)$

$N(Y) = \max\{n \in \mathbb{N}_0 | \sum\limits_{i=1}^{n} \leq Y\}$, wobei X_i *stid* $\sim Exp(\lambda)$

Bekannt: $P^{N(Y)|Y=t} = Poi(\lambda t)\ \forall t > 0$

$$\begin{aligned}
P(N(Y) = k) &= \int_0^\infty P(N(Y) = k | Y = t) \mu e^{-\mu t} = \int_0^\infty \frac{(\lambda t)^k}{k!} e^{-\lambda t} \mu e^{-\mu t} dt \\
&= \frac{\lambda^k \mu}{k!} \int_0^\infty t^k e^{-(\lambda+\mu)t} dt \qquad \text{(nun geschickt erweitern!)} \\
&= \frac{\lambda^k \mu}{k!} \frac{k!}{(\lambda+\mu)^{k+1}} \underbrace{\int_0^\infty \frac{(\lambda+\mu)^{k+1}}{k!} t^k e^{-(\lambda+\mu)t} dt}_{=1, \text{ da } \Gamma-\text{Dichte}} \\
&= \frac{\mu}{\lambda+\mu} \left(\frac{\lambda}{\lambda+\mu} \right)^k \qquad , k \in \mathbb{N}_0 \\
&= \left(1 - \frac{\mu}{\lambda+\mu} \right)^k \frac{\mu}{\lambda+\mu}
\end{aligned}$$

Also: $N(Y) \sim Geo\left(\frac{\mu}{\lambda+\mu} \right)$

8 Grenzwertsätze

Zur Motivation:

1. unendlicher Münzwurf, W'keit für Kopf = p (unbekannt)
   Frage: Ist die Münze fair? Welches p?
   Idee: Werfe die Münze sehr oft in unabhängigen Versuchen.
   $h_n = \frac{Anzahl\ der\ Würfe\ mit\ Kopf}{n=Anzahl\ aller\ Würfe}$
   Frage: Konvergiert h_n gegen p ($n \to \infty$)?

2. Wartesystem: Anzahl der in einem Zeitintervall ankommenden Kunden $X_i \sim Poi(\lambda)$. (Zeitintervall mit Länge 1)
   Anzahl der bis zur Zeit n ankommende Kunden: $\sum_{i=1}^{n} X_i \sim Poi(n\lambda)$
   Frage: Welche Verteilung ergibt sich asymptotisch (n groß)?
   $\exists$ Konstanten a_n, b_n : $\frac{Y_n - a_n}{b_n} \overset{as}{\sim} N(0,1)$
   Gilt die Aussage, wenn $X_i \sim Bin(k,p)$? Ja! Auch, wenn die Verteilung der X_i unbekannt.

Definition 8.1 $\{X_i\}_n \in \mathbb{N}$ sei Folge von ZV, X ZV auf W'raum $(\Omega, \mathfrak{A}, P)$.
Die Folge heißt

a) **P-fast-sicher konvergent** gegen X, wenn

$$P(\{\omega | \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = P(\lim_{n\to\infty} X_n = X) = 1$$

Bez.: $X_n \underset{n\to\infty}{\longrightarrow} X$ P-fs, $\lim\limits_{n\to\infty} X_n = X$ P-fs.

b) **P-stochastisch konvergent** gegen X, wenn

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0 \ \ \forall \epsilon > 0$$

Bez.: $X_n \underset{n\to\infty}{\longrightarrow} X$ P-stoch, $\lim\limits_{n\to\infty} X_n = X$ P-stoch.

c) **schwach- oder verteilungskonvergent** gegen X, wenn

$$\lim_{n\to\infty} F_n(x) = F(x) \qquad \text{f.a. Stetigkeitspunkte von } F$$

wobei $X_n \sim F_n, \ X \sim F$ (Grenzverteilungsfunktion)
Bez.: $X_n \overset{as}{\sim} X, \ X_n \overset{\mathfrak{D}}{\sim} X$

Lemma 8.2 $P(\limsup\limits_{n\to\infty}\{|X_n - X| > \epsilon\}) = 0 \Longleftrightarrow X_n \underset{n\to\infty}{\longrightarrow} X$ P-fs

Beweis: $''\Rightarrow''$ $\forall\epsilon = \frac{1}{k}:\ P(|X_n - X| > \frac{1}{k}$ für unendlich viele $n) = 0$
Es folgt: $P(\bigcup_k \{\omega \big| \, |X_n(\omega) - X(\omega)| > \frac{1}{k}$ für unendlich viele $n\}) \leq \sum_k P(...) = 0$
$P(\bigcup_k \{\omega|...) = \{\omega | \exists k \forall n_0 \exists n \geq n_0 : |X_n(\omega) - X(\omega)| > \frac{1}{k}\}$ Komplement liefert:
$P(\{\omega | \forall k \exists n_0 \forall n \geq n_0 : |X_n(\omega) - X(\omega)| \leq \frac{1}{k}\}) = 1$, d.h. $X_n \longrightarrow X$ P-fs

$''\Leftarrow''$ trivial

Satz 8.3 (Zusammenhänge zwischen Konvergenzarten)

a) $X_n \longrightarrow X$ P-fs $\overset{(i)}{\Longrightarrow}$ $X_n \longrightarrow X$ P-stoch $\overset{(ii)}{\Longrightarrow}$ $X_n \overset{\mathfrak{D}}{\longrightarrow} X$
(Umkehrungen sind i.A. falsch)

b) $\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty\ \ \forall \epsilon > 0 \Longrightarrow X_n \longrightarrow X$ P-fs
("Schnelle" stoch. Konvergenz impliziert fast sichere Konvergenz)

Beweis:

a) (i) $\forall\epsilon > 0$ gilt: $0 \leq \lim_{n\to\infty} P(|X_n - X| > \epsilon) \leq \lim_{n\to\infty} P(\bigcup_{m\geq n} \{|X_n - X| > \epsilon\})$
$\overset{Stetigkeit}{=} P(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|X_m - X| > \epsilon\}) = P(\limsup_{n\to\infty} |X_n - X| > \epsilon\})$
$= P(|X_n - X| > \epsilon$ für unendlich viele $n) = 0$, da $X_n \longrightarrow X$ P-fs (wg. Lemma 8.2)

(ii) Mathar & Pfeiffer, Lemma 2.3.2 Seite 137

b) Sei $\epsilon > 0$. Mit Borel-Cantelli-Lemma und Lemma 8.2:
$\sum_{N=1}^{\infty} P(|X_n - X| > \epsilon) \overset{B.C.L.}{\Longrightarrow} P(\limsup_{n\to\infty} \{|X_n - X| > \epsilon\}) = 0 \overset{8.2}{\Longrightarrow}$ Beh. $\square$

Lemma 8.4 (Tschebyscheff-Ungleichung)
X sei ZV mit $Var(X) < \infty$. Dann gilt $\forall\ \epsilon > 0$:

$$P(|X - EX| > \epsilon) \leq \frac{Var(X)}{\epsilon^2}$$

Beweis: Mit Markov-Ungleichung (6.5 d)): $P(|X| > c) \leq \frac{EX}{c}\ \forall c > 0$
$P(|X - EX| > \epsilon) = P((X - EX)^2 > \epsilon^2) \leq \frac{E(|X-EX|^2)}{\epsilon^2} = \frac{Var(X)}{\epsilon^2}$ $\square$

Satz 8.5 (Starkes Gesetz großer Zahlen, SGGZ)
$\{X_n\}\ n \in \mathbb{N}$ sei eine Folge von pw. unkorrelierten ZV auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit $Var(X_n) \leq M < \infty\ \forall n \in \mathbb{N}$.
Dann gilt:

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - E(X_i)) \underset{n\to\infty}{\longrightarrow} 0 \qquad \text{P-fs}$$

Beweis: O.B.d.A. $E(X_i) = 0 \ \ \forall i,\ X_i$ pw. unkorreliert. Setze $\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$.
Zeige $\forall \epsilon > 0 : \sum\limits_{i=1}^{\infty} P(|\overline{X}_i| > \epsilon) < \infty$. Dann gilt $\overline{X}_n \longrightarrow 0$ P-fs (8.3 b)
Betrachte zunächst die Folge der Quadrate: $\overline{X}_{n^2}$. Mit Tschebyscheff-Ungleichung folgt: $P(|\overline{X}_{n^2}| > \epsilon) \leq \frac{Var(\overline{X}_{n^2})}{\epsilon^2} \leq \frac{\frac{1}{n^4}n^2 M}{\epsilon^2} = \frac{M}{n^2\epsilon^2} \ \ \forall n \in \mathbb{N},\ \epsilon > 0$. Also:
$\sum\limits_{n=1}^{\infty} P(|\overline{X}_{n^2}| > \epsilon) \leq \sum\limits_{n=1}^{\infty} \frac{M}{n^2\epsilon^2} = \frac{M}{\epsilon^2}\sum\limits_{n=1}^{\infty}\frac{1}{n^2} < \infty \quad \Longrightarrow \quad \overline{X}_{n^2} \longrightarrow 0$ P-fs
Nun folgt der Nachweis für die Werte zwischen den Quadraten:
Dazu sei $n = n(k)$ def. durch $n^2 \leq k \leq (n+1)^2,\ k \in \mathbb{N}$.
$Var(k\overline{X}_k - n^2\overline{X}_{n^2}) = Var(\sum\limits_{i=n^2+1}^{k} X_i) \leq (k-n^2)M$

$$\boxed{\begin{aligned}&Var(k\overline{X}_k - n^2\overline{X}_{n^2}) \overset{Vor.}{=} Var(k\tfrac{1}{k}\sum_{i=1}^{k} X_i - n^2\tfrac{1}{n^2}\sum_{i=1}^{n^2} X_i) = Var(\sum_{i=1}^{k} X_i - \sum_{i=1}^{n^2} X_i)\\ &= Var(\sum_{i=n^2+1}^{k}) = \sum_{i=n^2+1}^{k} Var(X_i) \leq \sum_{i=n^2+1}^{k} M = (k-n^2)M\end{aligned}}$$

Es folgt (Tschebyscheff): $P(|k\overline{X}_k - n^2\overline{X}_{n^2}| \geq \epsilon n^2) \leq \frac{(k-n^2)M}{\epsilon^2 n^4}$ also mit Satz 8.3 b:
$\frac{k}{n(k)^2}\overline{X}_k - \overline{X}_{n(k)^2} \underset{k\to\infty}{\longrightarrow} 0$ P-fs
Insgesamt $\exists F_1 \in \mathfrak{A},\ P(F_1) = 1 \ \ \forall \omega \in \mathbb{F}_1 : \overline{X}_{n^2} \longrightarrow 0 \ \ (n \to \infty)$
$\exists F_2 \in \mathfrak{A},\ P(F_2) = 1 \ \ \forall \omega \in \mathbb{F}_2 : \frac{k}{n(k)^2}\overline{X}_k - \overline{X}_{n(k)^2} \longrightarrow 0 \ \ (k \to \infty)$
$\frac{k}{n(k)^2} \longrightarrow 1 \ \ (k \to \infty)$
Es folgt: $\overline{X}_n \longrightarrow 0 \ \ \forall \omega \in F_1 \cap F_2$, wobei:
$P(F_1 \cap F_2) = \underbrace{P(F_1)}_{1} + \underbrace{P(F_2)}_{1} - \underbrace{P(F_1 \cup F_2)}_{1} = 1$
D.h.: $\overline{X}_n \longrightarrow 0$ P-fs □

Bemerkung 8.6

a) Wegen Satz 8.3 (f.s. Konvergenz impliziert stochastische):
$\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - EX_i) \underset{n\to\infty}{\longrightarrow}$ P-fs $\quad\Longrightarrow\quad \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - EX_i) \underset{n\to\infty}{\longrightarrow} 0$ P-stoch
Die rechte Aussage heißt **schwaches Gesetz großer Zahlen**

b) Das SGGZ gilt auch, wenn $\{X_n\}_{n\in\mathbb{N}}$ stid und $E(X_1)$ existiert (ohne die Existenz der Varianz zu fordern). Dann gilt mit $\mu = E(X_i)$:
$\frac{1}{n}\sum\limits_{i=1}^{n} X_i \underset{n\to\infty}{\longrightarrow} \mu$ P-fs

Beweis: sehr aufwendig... □

Beispiel 8.7 (Anwendung des SGGZ)
$\{X_n\}_{n\in\mathbb{N}}$ sei Folge von stid ZV'en, $A \in \mathcal{B}^1,\ P(X_n \in A) = p \ \forall i \in \mathbb{N}$

$\sum\limits_{i=1}^{n} \mathbb{1}_A(X_i)$: Anzahl des Auftretens $\{X_i \in A\}$ bis zum n-ten Versuch.
$Y_i := \mathbb{1}_A(X_i)$ sind stid mit $E(Y_i) = P(X_i \in A) = p \quad \forall i \in \mathbb{N}$
Mit SGGZ: $\frac{1}{n}\sum\limits_{i=1}^{n} Y_i = \frac{1}{n}\underbrace{\sum\limits_{i=1}^{n} \mathbb{1}_A(X_i)}_{\text{rel. Häufigkeit des Auftretens von } \{X_i \in A\}} \longrightarrow p$ (für $n \to 0$) P-fs

Beispiel 8.8 $\{X_n\}_{n\in\mathbb{N}}$ sei stid Folge von ZV'en. $\mu = EX_i,\ \sigma^2 = Var(X_i)$. Dann $\frac{1}{n}\sum\limits_{i=1}^{n} X_i \underset{n\to\infty}{\longrightarrow} \mu$ P-fs (d.h. $\overline{X}_n$ ist **stark konsistenter Schätzer für** μ)
$\frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \underset{n\to\infty}{\longrightarrow} E(X_1^2)$ P-fs (wegen SGGZ; ersetze X_i durch X_i^2)
$\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 - (\overline{X}_n)^2 \longrightarrow E(X_1^2) - \mu^2 = Var(X_1) = \sigma^2$ (wegen SGGZ; d.h. $\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2$ ist ein **stark konsistenter Schätzer für** σ^2)

Satz 8.9 (Zentraler Grenzwertsatz, ZGWS)
$\{X_n\}_{n\in\mathbb{N}}$ sei stid Folge, $\mu = E(X_n),\ \sigma^2 = Var(X_n)$ existiere ($\sigma^2 > 0$)
Dann gilt

$$\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \overset{as}{\sim} N(0,1) \ \ (n \to \infty),$$

d.h. $P(\frac{1}{\sigma\sqrt{n}}\sum\limits_{i=1}^{n}(X_i - \mu) \leq z) \longrightarrow \Phi(z) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{z} e^{-\frac{x^2}{2}}dx \ \ \forall z \in \mathbb{R}$

Beweis: z.B. Casella & Berger 216 ff.

Für $S_n = \sum\limits_{i=1}^{n} X_i$ gilt $E(S_n) = n\mu,\ Var(S_n) = n\sigma^2$
$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{1}{\sigma\sqrt{n}}\sum\limits_{i=1}^{n}(X_i - \mu) \overset{as}{\sim} N(0,1) \ \ (n \to \infty)$
Dies bezeichnet man als die **standardisierte Summe**, und diese konvergiert gegen die Standardnormalverteilung ($N(0,1)$)

Beispiel 8.10 $\{X_n\}_{n\in\mathbb{N}}$ *stid* $\sim Bin(1,p),\ \ 0<p<1$. Es gilt: $E(X_n) = p,\ Var(X_n) = p(1-p)$
Mit ZGWS: $\frac{\sum\limits_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \overset{as}{\sim} N(0,1) \ \ (n \to \infty)$
Also:

$$P\left(\sum_{i=1}^{n} X_i \leq z\right) = \left[\sum_{k=0}^{\lfloor z\rfloor}\binom{n}{k}p^k(1-p)^{n-k} \ \ \text{(sehr aufwendig für große } n\text{)}\right]$$

$$
\begin{aligned}
&= P\left(\underbrace{\frac{\sum\limits_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}}}_{\overset{as}{\sim} N(0,1)} \le \frac{z-np}{\sqrt{np(1-p)}} \right) \\
&\approx \Phi\left(\frac{z-np}{\sqrt{np(1-p)}} \right), \ \textit{wobei}\ \Phi(z) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z} e^{-\frac{x^2}{2}} dx
\end{aligned}
$$

Bemerkenswert ist, daß die Konvergenz unabhängig davon ist, welche spezielle Gestalt die ursprüngliche Verteilung der Folge $\{X_n\}_{n\in\mathbb{N}}$ besitzt, und diese lediglich über die beiden ersten Momente Erwartungswert und Varianz als zentrierende Größen in die Konvergenz eingeht. Dies erklärt, warum man häufig Normalverteilungsannahmen in solchen stochastischen Modellen findet, in denen die Summen unabhängiger ZV eine Rolle spielen (siehe Mathar/Pfeifer, S. 143)

Allgemein gilt für eine Folge von *stid* ZV'en: ($\{X_n\}_{n\in\mathbb{N}}$ *stid*):

$$
\boxed{P\left(\sum_{i=1}^{n} X_i \le z\right) = P\left(\frac{\sum\limits_{i=1}^{n} X_i - n\cdot E(X_1)}{\sqrt{n\cdot Var(X_1)}} \le \frac{z-n\cdot E(X_1)}{\sqrt{n\cdot Var(X_1)}} \right) \overset{ZGWS}{\approx} \Phi\left(\frac{z-n\cdot E(X_1)}{\sqrt{n\cdot Var(X_1)}} \right)}
$$

(Approximation ist "brauchbar", falls $n \ge 30$)

Anwendung: Lieferung von n Teilen, Gut-Schlecht-Prüfung, W'keit, für Defekt: $p = 0,08$.
Gesucht: W'keit, daß Lieferung mehr als 9% defekt!

$$
\begin{aligned}
P\left(\sum_{i=1}^{n} X_i > 0,09n\right) &= 1 - P\left(\sum_{i=1}^{n} X_i \le 0,09n\right) \\
&\approx 1 - \Phi\left(\frac{0,09n - 0,08n}{\sqrt{n0,08(1-0,08)}} \right) \\
&= 1 - \Phi\left(\frac{0,01}{\sqrt{0,08\cdot 0,92}} \sqrt{n} \right) \\
&= 1 - \Phi\left(0,03686\sqrt{n}\right) \\
&= \ldots
\end{aligned}
$$

9 Schätzfunktionen und Konfidenzintervalle

Zur Motivation:

a) Werfe Münze in unabh. Versuchen, W'keit für Kopf= p (unbekannt). Schätze p. Modell: $X_1, ..., X_n$ *stid* $Bin(1, p)$.
$\widehat{p} = \widehat{p}(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \longrightarrow \quad p$ P-fs (bekannt, SGGZ) ist ein vernünftiger Schätzer für p.
Beachte: $\widehat{p}(X_1, ..., X_n)$ ist eine ZV. Setzt man **Realisationen** $x_1, ..., x_n$ ein, so erhält man den **Schätzwert** $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

b) $X_1, ..., X_n$ *stid* $Exp(\lambda)$. Familie der $Exp-$Verteilung mit Parameter λ, parametrische Familie.
$\widehat{\vartheta} = \frac{1}{n}\sum_{i=1}^{n} \quad \longrightarrow \quad \frac{1}{\lambda} = E(X_1)$ P-fs (SGGZ)
$\widehat{\vartheta}$ ist ein Schätzer für $g(\lambda)?\frac{1}{\lambda} = E(X_1)$

c) In a) und b) wird der Erwartungswert durch das arithmetische Mittel geschätzt. Auch schwierigere Situationen, z.B.:
M: Anzahl der regelmäßigen Besucher einer Web-Seite, unbekannt. Schätze M! Wie? Intuitives Verfahren:

- Merke die Adressen von n Besuchern (Log-File), 'markiere' Besucher
- Zu späterem Zeitpunkt: Merke m Adressen, bestimme Anzahl x von bereits markierten Besuchern.

Unter entsprechenden Voraussetzungen sollte $\frac{n}{M} \approx \frac{x}{m}$ (ungefähr gleich).
Ein vernünftiger Schätzer für $M = \lfloor \frac{nm}{x} \rfloor$. Kann dieses Vorgehen exakt begründet werden?

9.1 Methoden zur Bestimmung von Schätzern

Definition 9.1 $X_i, ..., X_n$ seien Zufallsvariablen (ZV) mit gemeinsamer Verteilung $P_{\vartheta}^{(X_1,...,X_n)}$, $\vartheta \in \Theta$ (Parameterraum). $g(\vartheta : \Theta \to \mathcal{Y})$ sei eine Funktion. Jede Abbildung $h(X_1, ..., X_n)$ mit Werten in $\mathcal{Y}$ heißt **statistische Schätzfunktion** oder (**Punkt-**)**Schätzer**.

Beispiel 9.2

a) $X_1, ..., X_n$ *stid* $\sim N(\mu, \sigma^2)$, $\vartheta = (\mu, \sigma^2)$, $\Theta = \mathbb{R} \times \mathbb{R}^+$,
$$g_1(\vartheta) = g_1((\mu, \sigma^2)) = \mu, \quad h_1(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{OK}$$
$$g_2(\vartheta) = g_2((\mu, \sigma^2)) = \sigma^2, \quad h_2(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad \text{OK}$$

b) $X_1, ..., X_n$ *stid* $\sim Exp(\lambda),\ \vartheta = \lambda,\ \Theta = \mathbb{R}^+$

$$g_1(\lambda) = \frac{1}{\lambda},\quad h_1(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{OK}$$

$$g_2(\lambda) \qquad h_2(X_1, ..., X_n) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^{-1} \qquad \text{vernünftig?}$$

Im Folgenden Methoden, um geeignete Schätzer zu finden.

Definition 9.3 (Maximum-Likelihood-Schätzer)
$f(x_1, ..., x_n|\vartheta),\ \vartheta \in \Theta$, sei eine (Zähl-)Dichte. $L(\vartheta|x_1, ...x_n) = f(x_1, ..., x_n|\vartheta)$ (als Funktion von ϑ bei gegebenen x_i) heißt **Likelihood-Funktion**.
$\widehat{\vartheta}(x_1, ...x_n)$ heißt **Maximum-Likelihood-Schätzer** (MLS), falls

$$L(\widehat{\vartheta}(x_1, ...x_n)|x_1, ..., x_n) = \sup_{\vartheta\in\Theta} L(\vartheta|x_1, ..., x_n) \quad \forall\ x_1, ..., x_n$$

$g(\widehat{\vartheta})$ heißt MLS von $g(\vartheta)$.

Konzept: $\widehat{\vartheta}(x_1, .., x_n)$ ist bei gegebenen Beobachtungen $x_1, ..., x_n$ der 'passendste' Parameter, der die Beobachtungen am wahrscheinlichsten macht.

Bemerkung

a) $\widehat{\vartheta}$ zu bestimmen, ist eine Maximierungsaufgabe, die oft durch Differenzieren (NST der 1. Abl.) gelöst werden kann.
b) Oft ist es günstiger, statt $L(\vartheta|x)$ die **Log-Likelihood-Funktion** $logL(\vartheta|x)$ zu betrachten (siehe folgendes Beispiel). Das Maximum von $L(\vartheta|x)$ und $logL(\vartheta|x)$ liegt an derselben Stelle.

Beispiel 9.4

a) $X_1, ..., X_n$ *stid* $\sim N(\mu, \sigma^2)$. Schätze $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$.

$$f(x_1, .., x_n|\vartheta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} = L(\vartheta|x_1, ..., x_n)$$

$$\log L(\vartheta|x_1, ..., x_n) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Setze $\tau = \sigma^2$, partielle Ableitung $= 0$ setzen, auflösen nach μ, τ :

(i) $\frac{\partial \log L}{\partial \mu} = +\frac{2}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \overset{!}{=} 0$

(ii) $\frac{\partial \log L}{\partial \tau} = -\frac{n}{2}\frac{2\pi}{2\pi\tau} + \frac{1}{2\tau^2}\sum_{i=1}^{n}(x_i-\mu)^2 \overset{!}{=} 0$

Aus (i) : $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$

(ii) : $-\frac{n}{2}\frac{1}{\tau} + \frac{1}{2\tau^2}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 0 \iff \tau = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$

$\Longrightarrow \widehat{\tau} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$

Insgesamt: $(\widehat{\mu}, \widehat{\sigma^2}) = (\overline{x}, \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2)$ ist der MLS für (μ, σ^2)

Bleibt noch zu prüfen, ob dies auch wirklich ein Maximum ist

b) $X_1, ..., X_n$ *stid* $\sim Exp(\lambda)$. Schätze $\vartheta = \lambda \in \mathbb{R}^+$

$f(x_1, ..., x_n|\vartheta) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} = L(\lambda|x_1, ..., x_n),\ x_i \geq 0$

$\log L(\lambda|x_1, ..., x_n) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i$ (Aufgabe: Maximiere nach λ)

$\frac{\partial \log L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \stackrel{!}{=} 0 \iff \frac{n}{\lambda} = \sum_{i=1}^{n} x_i \iff \lambda = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}$

Also: $\widehat{\lambda} = \frac{1}{\overline{x}}$ ist ein MLS für λ.

(Noch z.z., $\widehat{\lambda}$ <u>ist</u> ein Max. von $L(\lambda|x_1, ..., x_n)$)

c) $X_1, ..., X_n$ *stid* $\sim Bin(1, p)$. Schätze $p \in [0, 1]$

$f(x_1, ..., x_n|p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} = L(p|x_1, ..., x_n)$

wobei die $x_i \in \{0, 1\}$

$\log L(p|x_1, ..., x_n) = \left(\sum_{i=1}^{n} x_i\right) \log p + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-p)$

(Maximiere über $p \in [0, 1]$!)

$\frac{\partial \log L}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\left(n - \sum_{i=1}^{n} x_i\right) \stackrel{!}{=} 0 \iff \frac{1}{p}\overline{x} - \frac{1}{1-p}(1-\overline{x}) = 0 \iff$

$p = \overline{x}$

Also: $\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$ ist ein MLS.

Eine weitere Methode zur Konstruktion von Schätzern: **Bayes-Methode**

1. Modelliere Vorkenntnisse über den Parameter ϑ durch eine Wahrscheinlichkeitsverteilung über Θ, die **a-priori-Verteilung**, beschrieben durch die (Zähl-)Dichte $\pi(\vartheta)$.

2. $f(x|\vartheta),\ x = (x_1, ..., x_n)$, sei die Zähldichte oder Dichte der Verteilung der ZV $X_1..., X_n$ bei Vorliegen von ϑ, aufgefaßt als bedingte Verteilung von $(X_1, ..., X_n)$, gegeben ϑ.

3. Gegeben die Beobachtungen $x = (x_1, ..., x_n)$. Die zu $f(\vartheta|x)$ gehörige Verteilung heißt **a-posteriori-Verteilung** von ϑ. $f(\vartheta|x)$ reflektiert den Kenntnistand über ϑ nach Beobachtungen von $x_1, ..., x_n$.

Bayes Verfahren
Gegeben $f(x|\vartheta) : \pi(\vartheta)$ a-priori-Verteilung. Berechnung von $f(\vartheta|x)$ wie folgt (hier im Falle von Dichten): $f(x) = \int f(x|\vartheta)\pi(\vartheta)d\vartheta$ (siehe Kapitel 7).
Also:

$$f(\vartheta|x) = \frac{f(x,\vartheta)}{f(x)} = \frac{f(x|\vartheta)\pi(\vartheta)}{\int f(x|\vartheta)\pi(\vartheta)d\vartheta}$$

Dabei tritt der Nenner als normierende Konstante auf.

Ein natürlicher Schätzer für ϑ

Definition 9.5 Bez. wie oben. Die ZV $\widehat{\vartheta}(x)$ besitzt die (Zähl-)Dichte $f(\vartheta|x)$. Dann heißt $E(\vartheta(x))$ **Bayes Schätzer** von ϑ.

Beispiel 9.6 Sei $X \sim Bin(n,p)$, n fest.
A-priori Verteilung für p sei $Beta(\alpha, \beta)$, $\alpha,\ \beta > 0$, mit Dichte:

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \quad 0 \leq p \leq 1$$

Es gilt: $X \sim Beta(\alpha, \beta) \implies E(X) = \frac{\alpha}{\alpha+\beta}$.
Sei $x \in \{0, ..., n\}$:

$$\begin{aligned} f(p|x) &= \frac{f(x|p)\pi(p)}{\int f(x|p)\pi(p)dp} = \frac{\binom{n}{x} p^x (1-p)^{n-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}}{\int f(x|p)\pi(p)dp} \\ &= \frac{\binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{x+\alpha-1}(1-p)^{n-x+\beta-1}}{\int f(x|p)\pi(p)dp} \\ &= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)} p^{x+\alpha-1}(1-p)^{n-x+\beta-1} \end{aligned}$$

Diese Lösung ergibt sich dadurch, daß eine Dichte herauskommen muß, und da nur die beden Teile mit dem Exponentialanteil variabel sind, ergibt sich eine *Beta*-Dichte mit Parametern $x + \alpha,\ n - x + \beta$.

Also: $\widehat{p}(x) = \frac{x+\alpha}{n+\alpha+\beta}$ ist Bayes-Schätzer zur a-priori Verteilung π.
Speziell: $\alpha = \beta = 1$ (dann $Beta(\alpha, \beta) = R(0, 1)$, Rechteck)

$\widehat{p}(x) = \frac{x+1}{n+2}$
MLS für p lautet: $\widehat{p}_{ML}(x) = \frac{x}{n}$

9.2 Gütekriterien für Schätzer

Definition 9.7 $H = h(X_1, ..., X_n)$ sei ein Schätzer für $g(\vartheta) \in \mathbb{R}$.
$E_\vartheta(H - g(\vartheta))^2 = E_\vartheta(h(X_1, ..., X_n) - g(\vartheta))^2$ heißt **mittlerer quadratischer Fehler** (MSE = mean squared error)

MSE mißt die mittlere quadratische Abweichung vom zu schätzenden Wert $g(\vartheta)$.
MSE ist eine Funktion von ϑ.
Es gilt:

$$\begin{aligned}
&E_\vartheta(H - g(\vartheta))^2 \\
&\quad = E_\vartheta(H - E_\vartheta H + E_\vartheta H - g(\vartheta))^2 \\
&\quad = E_\vartheta(H - E_\vartheta H)^2 + 2(E_\vartheta H - g(\vartheta))\underbrace{(E_\vartheta(H - E_\vartheta H))}_{=0} + (E_\vartheta H - g(\vartheta))^2 \\
&\quad = E_\vartheta(H - E_\vartheta H)^2 + (E_\vartheta H - g(\vartheta))^2 \\
&\quad = \underbrace{Var_\vartheta(H)}_{\text{Präzision, Variabilität des Schätzers}} + \underbrace{(Bias_\vartheta(H))^2}_{Schiefe,\ Genauigkeit\ des\ Schätzers}
\end{aligned}$$

Wichtig ist der Fall, daß $Bias_\vartheta(H) = 0$ ist.

Definition 9.8 Ein Schätzer H heißt **erwartungstreu** (unbiased) für $g(\vartheta)$, wenn $Bias_\vartheta(H) = 0 \ \forall \vartheta \in \Theta$, falls also $E_\vartheta(H) = g(\vartheta) \ \ \forall \vartheta \in \Theta$

Offensichtlich gilt für E-treue Schätzer. MSE $(H): \ E_\vartheta(H - g(\vartheta))^2 = Var_\vartheta(H)$

Beispiel 9.9 $X_1, ..., X_n \sim N(\mu, \sigma^2)$ $(stid)$ $\quad \left[\overline{X}, \ \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2 \quad \text{E} - \text{treu?}\right]$

a) $\mu = H(X_1, ..., X_n) = \frac{1}{n}\sum\limits_{i=1}^{n} X_i = \overline{X}$

$E_\vartheta(\overline{X}) = E_\vartheta(\frac{1}{n}\sum\limits_{i=1}^{n} X_i) = \frac{1}{n}E_\vartheta(\sum\limits_{i=1}^{n} X_i) = \frac{1}{n}\sum\limits_{i=1}^{n} E_\vartheta(X_i) = \frac{1}{n}\sum\limits_{i=1}^{n} \mu = \mu$

Also: $\overline{X}$ ist erwartungstreu für μ.

b) $\frac{1}{\sigma^2} = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$

Vorüberlegungen:

$Y \sim N(0,1) \quad \Longrightarrow \quad E(Y^2) = 1 \quad$ (Bsp. 6.13 c))

$Y \sim N(0,1) \quad \Longrightarrow \quad Z = \sigma Y + \mu \sim N(\mu, \sigma^2)$

$E(Z^2) = E(\sigma Y + \mu)^2 = \sigma^2 \underbrace{E(Y^2)}_{=1} + 2\sigma\mu \underbrace{E(Y)}_{=0} + \mu^2 = \sigma^2 + \mu^2$

$\widehat{\sigma}^2 = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 - \overline{X}^2$

$$E_\vartheta(\widehat{\sigma}^2) = \frac{1}{n}\sum_{i=1}^{n}\underbrace{E(X_i^2)}_{\sigma^2+\mu^2} - \underbrace{E(\overline{X}^2)}_{\frac{\sigma^2}{n}+\mu^2,\text{ da }\overline{X}\sim N(\mu,\frac{\sigma^2}{n})} = \sigma^2+\mu^2-\frac{\sigma^2}{n}-\mu^2 = \frac{n-1}{n}\sigma^2$$

Also: $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ ist nicht erwartungstreu für σ^2.

Aber: $\widehat{\sigma}^2$ ist **asymptotisch erwartungstreu**, d.h. $\lim_{n\to\infty} E(\widehat{\sigma}^2) = \sigma^2$

Aber: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ ist erwartungstreu für σ^2.

$\boxed{\text{Der Vorfaktor ergibt sich aus } \frac{\frac{1}{n}}{\frac{n-1}{n}}}$

Denn: $E(S^2) = E\left(\frac{n}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right) = \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 = \sigma^2$

Beispiel 9.10 (Gütevergleich von 2 Schätzern)
Sei $X \sim Bin(n,p)$ und folgende 2 Schätzer sind gegeben:

- MLS $\widehat{p}_{ML} = \frac{x}{n}$
- Bayes-Schätzer zur a-priori Verteilung $Beta(\alpha,\beta)$ $\widehat{p}_B = \frac{x+\alpha}{n+\alpha+\beta}$ (Beisp. 9.6)

Welcher ist nun besser? Vergleich mit MSE:

- $E_p(\widehat{p}_{ML}-p)^2 = Var(\widehat{p}_{ML}) = Var(\frac{1}{n}X) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$
- $E_p(\widehat{p}_B-p)^2 = Var(\widehat{p}_B)+Bias_p(\widehat{p}_B)^2 = Var\left(\frac{x+\alpha}{n+\alpha+\beta}\right)+\left(E_p\left(\frac{x+\alpha}{n+\alpha+\beta}\right)-p\right)^2$
  $= \frac{np(1-p)}{(n+\alpha+\beta)^2}+(\frac{np+\alpha}{n+\alpha+\beta}-p)^2$

Wähle nun α,β so, daß MSE($\widehat{p}_B$) konstant, d.h. kein Wert von p wird bei der Schätzung bevorzugt!

Lösung: $\alpha=\beta=\frac{1}{2}\sqrt{n}$ (setze oben ein, dann fällt das p heraus)
Dann gilt:

$$\widehat{p}_B = \frac{x+\frac{1}{2}\sqrt{n}}{n+\sqrt{n}}$$

$$E_p((\widehat{p}_B-p)^2 = \frac{n}{4(n+\sqrt{n})^2}$$

9.3 Konfidenzintervalle

Punktschätzer: Bestimme $\widehat{\vartheta}$, das möglichst nahe am Parameter ϑ liegt.
Ziel jetzt: Angabe von Schranken, so daß der Parameter ϑ mit vorgegebener W'keit innerhalb der Schranken liegt.

Idee: $X_1, ..., X_n$ $stid \sim Bin(1,p)$. Mit schwachem GGZ:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > \epsilon\right) \underset{n\to\infty}{\longrightarrow} 0 \quad \forall \epsilon > 0$$

$$\iff \quad P(\overline{X} - \epsilon < p < \overline{X} + \epsilon) \underset{n\to\infty}{\longrightarrow} 1$$

Für große n wird der Parameter p durch das zufällige Intervall $[\overline{X} - \epsilon, \overline{X} + \epsilon]$ mit hoher W'keit überdeckt. Nutze Kenntnisse über die Verteilung von $\overline{X}$, um ϵ, $P(...)$ zu quantifizieren.

Definition 9.11 $X_1, ..., X_n$ seien ZV mit gemeinsamer Verteilung $P^{X_1,...,X_n}$. Ein Intervall der Form $[L(X_1, ..., X_n), U(X_1, ..., X_n)]$ heißt **Konfidenzintervall zum Niveau** $1 - \alpha$ für $g(\vartheta)$, falls

$$P_\vartheta(g(\vartheta) \in [L(X_1, ..., X_n), U(X_1, ..., X_n)]) \geq 1 - \alpha \quad \forall \vartheta \in \Theta$$

Beispiel 9.12 $X_1, ..., X_n$ $stid \sim N(\mu, \sigma^2)$ (σ^2 fest)
Dann $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ und $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
Bestimme nun n so, daß $P\left(\left|\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq u\right) = 1 - \alpha \quad (\star)$
Wähle $u = u_{1-\frac{\alpha}{2}}$ als $\left(1 - \frac{\alpha}{2}\right)$-Fraktil der N(0,1)-Verteilung. Nun auflösen von $(\star)$ nach μ:
$P\left(|\overline{X} - \mu| \leq \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}\right) = P\left(\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$
Also: $\left[\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}, \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}\right]$ ist ein $(1 - \alpha)$-Konfidenzintervall für μ.

<u>Beachte</u> Die Länge des Konfidenzintervalls
- fällt mit wachsendem n
- wächst mit wachsendem Niveau $1 - \alpha$

Auch einseitige Konfidenzintervall: $P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq u\right) = 1 - \alpha$
Wähle $u = u_{1-\alpha}$, dann: $P\left(\overline{X} - \mu \leq \frac{\sigma}{\sqrt{\mu}} u_{1-\alpha}\right) = P\left(\mu \geq \overline{X} - \frac{\sigma}{\sqrt{\mu}} u_{1-\alpha}\right) = 1-\alpha$
Also: $\left[\overline{X} - u_{1-\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right]$ ist **einseitiges** $(1 - \alpha)$**-Konfidenzintervall für** μ

All diese Berechnungen gelten für **festes** σ^2. Für variable Varianzen siehe Literatur über W'keitsrechnung!!!

Beispiel 9.13 (Approx. Konfidenzintervalle für W'keiten)
$X_1, ...X_n \sim Bin(1,p)$. Dann $\frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{as}{\sim} N(0,1)$ (Bsp. 8.10)
Es gilt: $P\left(\frac{|\overline{X}-p|}{\sqrt{\frac{p(1-p)}{n}}} \leq u_{1-\frac{\alpha}{2}}\right) \approx 1-\alpha$, wobei $u_{1-\frac{\alpha}{2}}$ Fraktil der $N(0,1)$-Verteilung ist.

$$|\overline{X}-p| = u_{1-\frac{n}{2}}\sqrt{\frac{p(1-p)}{n}} \quad \Longleftrightarrow \quad (\overline{X}-p)^2 = u^2_{1-\frac{\alpha}{2}}\frac{p(1-p)}{n}$$

ist eine quadratische Gleichung in p mit Lösung $p_L(\overline{X}) \leq p_U(\overline{X})$.

$$\left[p_L(\overline{X}), p_U(\overline{X})\right]$$

ist ein **approximatives** $(1-\alpha)$**-Konfidenzintervall für** p.

ENDE DER VORLESUNG

Index